BAZE TEORETICE ALE P S I H O D I A G N O S T I C U L U I

AUREL ION CLINCIU

UNIVERSITATEA TRANSILVANIA BRAŞOV

2009

CAPITOLUL 1

INTRODUCERE

1.1. Psihodiagnoza, domeniu de vocaţie al profesiunii de psiholog

O revistă literaturii privind principala activitate a profesiunii de psiholog clinician scotea în evidenţă, acum două decenii, un întreg spectru de preocupări, dintre care *evaluarea* şi *psihodiagnoza* deţineau cea mai mare pondere (50%), urmate de *psihoterapie* (30%), de *cercetarea ştiinţifică* (30%), de *consiliere* (5%) şi de *terapia de mediu şi socioterapie* (5%) (vezi Ionescu, 1985, p. 51).

Plasarea în top a *psihodiagnozei* nu este specifică numai domeniului clinic, ci întregului teritoriu al psihologiei aplicate, pentru că, prin instrumentele ei înalt standardizate, ea oferă o cunoaştere de adâncime asupra fiinţei umane, pe care se pot fundamenta ulterior consilierea, psihoterapia sau ghidarea vocaţională. Afirmaţia lui Perse (1972, p. 112), conform căreia „la ora actuală, examenul psihometric constituie partea cea mai importantă a intervenţiei psihologului (…) metoda testelor fiind poate singura datorită căreia specificitatea şi autonomia sa nu sunt contestate" îşi păstrează şi astăzi valabilitatea, cu menţiunea că termenul de testare ar trebui completat cu acela mai cuprinzător de *evaluare*.

Împărtăşim punctul de vedere al lui Kaplan şi Saccuzzo (1993, p. iii) care îşi exprimă „mâhnirea pentru că prea mulţi studenţi din colegii văd cursul de testare psihologică ca plictisitor şi nelegat de obiectivele lor sau de interesele lor pentru carieră". Însăşi evoluţia psihologiei, prin ruperea ei din trunchiul disciplinelor speculativ-filosofice spre cele practic-aplicative, este marcată de intrarea ei în laborator în 1879, prin Wundt şi de supunerea la rigoarea ştiinţifică a experimentului, tendinţă continuată de Galton, cu al său laborator de antropometrie din 1884, sau de James McKeen Cattell care, în 1888, propunea termenul de *teste mintale*, ce urmau să fie aplicate în laboratorul său de testare.

Recursul la experiment a fost indisolubil legat de apelul la cifră, la cuantificare şi măsurătoare, care făceau ca psihologia să îşi dezvolte instrumente precise de investigaţie, metode din ce în ce mai elaborate de analiză a datelor rezultate, ce confereau cunoaşterii psihologice rigoare, exactitate şi generalitate. Este deci cu totul surprinzător faptul că acest curs de evaluare-testare psihologică, care pregăteşte studentul pentru ceea ce va fi partea fundamentală a profesiunii sale, să fie găsit plictisitor, ceea ce s-ar putea explica eventual prin faptul că aceste domeniu de supraspecializare a devenit din ce în ce mai tehnic, încorporând procedee de lucru şi proceduri statistice sofisticate, care de multe ori estompează frumuseţea

lui provocatoare. Acest impediment nu poate fi depăşit decât prin faptul ca profesorul să fie el însuşi un practician cu multă experienţă, astfel încât printre rânduri să se simtă freamătul problemelor vii, al implicării profunde şi devotate într-o profesie care a devenit, prin toate acestea, o *vocaţie.*

1.2. Psihodiagnoză, testare şi evaluare

Psihodiagnoza este un concept foarte puţin sau deloc utilizat de şcoala anglo-saxonă şi aceasta probabil din cauza nedoritei apropieri a acestui termen de acela de *diagnostic,* creat şi consacrat de practica medicală. În acest spaţiu cultural a fost folosit extensiv termenul de testare (*testing*), pentru a circumscrie teritoriul cunoaşterii exacte şi obiective, de tip psihometric a persoanei, deşi în ultimele decenii cel de evaluare (*assessment*) a câştigat tot mai mult teren. Şi aceasta din mai multe motive:

- Chiar dacă din capul locului testele au fost destinate unor scopuri nu doar constatative sau exclusiv ştiinţifice (de cunoaştere), ci şi-au propus să aducă o schimbare benefică în viaţa oamenilor, ele au generat o anumită aprehensiune, rezistenţă sau chiar o reacţie adversă. Deoarece testele şi itemii ce le compun au fost păstraţi departe de „ochiul public", ca de altfel şi procedurile de scorare, interpretare şi utilizare a rezultatelor, ele au creat un val de mister ce a generat teamă în rândul publicului larg, cu atât mai mult cu cât utilizarea lor a cunoscut o adevărată explozie între cele două războaie mondiale.
- Psihologii s-au văzut constrânşi de reglementările legale din domeniu, de teama publicului larg faţă de teste (şi de consecinţele acesteia) să privească cu interes şi spre alte instrumente ale profesiunii lor (interviul, ancheta, studiul de caz etc.), testul rămânând instrumentul cel mai elaborat, în mâna examinatorului cu o foarte înaltă calificare, la care are acces după stagii riguroase de pregătire, asistată, controlată şi certificată de specialişti recunoscuţi.
- Evaluarea psihologică (*assessment*) a devenit termenul integrator, el fiind conceput ca o rezolvare de probleme, deoarece fiecare variabilă din natură este rezultatul combinării unor factori diferiţi, de unde necesitatea unor metode variate (printre care şi testele), depinzând de obiectivele, scopurile sau circumstanţele în care se desfăşoară procesul evaluativ.
- Testarea psihologică a devenit specifică unor scopuri mai înguste şi mai bine circumscrise, în timp ce evaluarea este mai largă, predicţiile şi deciziile ei bazându-se pe o multitudine de surse.

- Ca şi obiective, testarea nu îşi propune să răspundă la întrebarea cum procesează persoana o anumită informaţie, fiind centrată pe produs (rezultatul la test) şi nu pe proces (maniera de a ajunge la acel rezultat).
- Evaluarea (inclusiv cea formativă) încearcă să înţeleagă o problemă în cadrul ei ecologic mai larg, de unde şi spectrul mult mai extins de informaţii utilizate (observaţia naturalistă, vizita la domiciliu, ancheta, interviul sau simpla discuţie cu părinţii, profesorii şi reprezentanţii autorităţii). Ea nu dezavuează procedurile standardizate, tipice testării, dar fiind conştientă că acestea nu pot răspunde la toate întrebările, valorizează şi informaţiile colaterale (stil de viaţă, atmosferă familială, interacţiunile din grup etc.).
- *Relaţia* cu clientul este diferită în testare, unde poate să existe o legătură discontinuă, multiplu mediată sau chiar absentă (în testarea de grup) sau o mai mică responsabilitate legală a examinatorului faţă de cel examinat. În evaluare relaţia este una de faţă-în-faţă, cu responsabilităţi explicite din partea examinatorului de a desfăşura o activitate în folosul clientului său.
- În practică, distincţia nu este aşa de tranşantă, aşa cum arată Cohen, Swerdlik şi Smith (1992), termenii fiind deseori utilizaţi ca echivalenţi şi interşanjabili, deşi distincţia dintre ei trebuie păstrată. Iată cum definesc autorii citaţi evaluarea psihologică (*assessment*): „proces de strângere şi de integrare a datelor psihologice, corelate scopului de a face o evaluare psihologică, utilizând instrumente ca testele, interviul, studiul de caz, observaţia comportamentală şi aparate special destinate, împreună cu procedee de măsurătoare adecvate" (p. 11).

Deoarece în terminologia românească nu s-a impus termenul de testare şi nici cel de evaluare, pentru circumscrierea domeniului termenul consacrat fiind cel de *psihodiagnostic,* ca şi în unele dintre ţările francofone, vom face distincţia dintre diagnosticul medical şi cel psihologic. După cum se ştie, medicul identifică o anumită boală, prin mijloace clinice (observaţie, anamneză, interviu clinic, examen somatic sau psihic etc.) şi paraclinice (analize de laborator, radiografie, echografie etc.) pentru a o introduce într-un cadru nosologic şi a formula un tratament. Schema consacrată de mai bine de un secol pentru multe din domeniile medicinei este: diagnostic *etiologic* (vizând cauza bolii), *patogenetic* (mecanismele ei de producere), diagnosticul *pozitiv* (raportarea semnelor la un sindrom) şi *diferenţial* (raportarea la entităţi apropiate din cadrul nosografic mai general). Toate acestea nu numai pentru a circumscrie boala, ci pentru a stabili un *prognostic* şi un *tratament* adecvat.

În psihologie, diagnosticul utilizează de asemenea o multitudine de mijloace, dintre care doar unele standardizate: testul, interviul, studiul de caz, observaţia comportamentului, tehnici derivate din metoda experimentală, ca reacţia electrogalvanică, reflexul fotopupilar sau timpul de reacţie ori cel de latenţă.

Sursele majore ale informaţiilor sunt analiza comportamentelor, a activităţii şi a produselor acestora, studiul expresiei şi al conduitei verbale (Şchiopu, 1976). Marea diferenţă în raport cu diagnosticul medical este aceea că în psihodiagnoză se are în vedere terenul psihologic nealterat al omului de pe versantul normalităţii (excepţie făcând munca psihologului din clinica neuropsihiatrică), pentru a face bilanţul caracteristicilor psihice definitorii.

În viziunea Ursulei Şchiopu, psihodiagnoza este „un act de analiză psihologică a unui caz şi, concomitent, o sinteză logică ce permite organizarea conştientă a măsurilor generale de influenţare a dezvoltării personalităţii cazului considerat" (op. cit., p. 33). Elementul comun este aşadar succesiunea dintre latura constatativă (diagnostică) şi cea prognostic-ameliorativă şi formativă (personalitatea optimală, selecţia omului potrivit pentru locul potrivit, personalizarea intervenţiei etc.). În acest sens psihodiagnoza este elementul de bază al psihologiei diferenţiale, pe care a şi făcut-o posibilă.

Deşi ocupă un teritoriu extrem de important al psihologiei aplicate, din care s-a născut şi pe care a dezvoltat-o continuu, psihodiagnoza nu se întâlneşte definită ca atare în dicţionarele de psihologie, după cum nici noţiunile de testare şi de evaluare, care o circumscriu, nu sunt definite. Eventual ele reţin accepţiunea iniţială a termenului propus de către Herman Rorschach, după care psihodiagnosticul „este o metodă de explorare a personalităţii, bazată pe interpretarea liberă de forme accidentale", cum ar fi petele de cerneală, ca în *Dicţionarul de psihologie Larousse*, al lui N. Sillamy (1996, p. 244), sau ca în *Dicţionarul de psihologie* al lui Doron şi Parot (1999, p. 624). Niciunul dintre ele nu defineşte termenul de diagnostic psihologic, dat ca şi titlu unei lucrări de referinţă a lui Richard Meili (*Manual psihodiagnostic*, 1964), dar nici pe aceia de testare-evaluare, în accepţiunea americană a termenilor utilizaţi de numeroşi autori (Anastasi, Cronbach, Gregory, Aiken, Kaplan, Saccuzzo etc.), pentru a-şi numi lucrările lor.

În *Dicţionarul encilcopedic de psihologie* din 1997, coordonat de Ursula Şchiopu, reţinem definiţia termenului care exprimă chiar viziunea autoarei, după care „diagnoza psihică se referă la omul viu şi concret şi constă dintr-o cunoaştere complexă a acestuia prin tehnici psihologice, aplicate conform unei strategii, ce permit colectări de informaţii organizate privind persoana dată" (op. cit., p. 542). Termenul de psihodiagnostic este prezent

şi la Mariana Roşca (*Metode de psihodiagnostic,* 1974) şi la alţi autori clujeni care, în deceniul al optulea au editat un valoros *Îndrumător psihodiagnostic*, în mai multe volume. Acesta a fost un foarte preţios ghid pentru psihologul practician, căci conţinea nu numai elemente de natură teoretică despre exigenţele examinărilor psihologice, ci prezenta o diversitate de instrumente practice, adaptate şi etalonate pe populaţie românească. Lucrare remarcabilă prin preocuparea ei de a oferi elemente de sprijin practic muncii psihologului a fost *Aptitudinile şi măsurarea lor*, a lui Bontilă, editată de Centrul de documentare şi publicaţii al Ministerului Muncii în 1971.

Lucrările mai noi, de după revoluţia din 1989, par a evita însă termenul de psihodiagnostic, ele vorbind de *Construirea şi utilizarea testelor psihologice* (Albu, 1998), despre *Cunoaşterea psihologică a persoanei* (Havârneanu, 2000), despre *Evaluarea în actul educaţional – terapeutic* (Vlad, 1999) sau despre *Testarea psihologică a copilului mic* (Mitrofan, 1997). Doar Valentina Horghidan mai conservă un concept care pare să devină inactual, cel de psihodiagnostic, în lucrarea intitulată chiar *Metode de psihodiagnostic* (1997). În aceasta autoarea include o multitudine de metode, pe lângă clasicele test şi chestionar apărând şi observaţia, convorbirea, ancheta, experimentul, scalele de evaluare a abilităţilor, metoda aprecierii obiective a personalităţii, fişa de caracterizare psihopedagogică a elevului şi chiar metodele sociometrice. Se produce astfel o lărgire exagerat de mare a cadrului metodologic al acestei discipline ce tinde să-şi subsumeze prin înglobare întreaga metodologie a ştiinţei psihologice, cu deschidere şi spre sociologie.

Opinăm că termenul tradiţional de psihodiagnostic poate fi păstrat, pentru că el include şi pe cel foarte tehnic şi consacrat de testare (*testing*), destinat determinărilor cantitative, dar şi pe cel de evaluare (*assessment*), desemnat să completeze şi să nuanţeze diagnosticul cu elemente calitative. Dar, pentru că testul psihologic rămâne instrumentul său cel mai tehnic şi cel mai puternic, cu o evoluţie neegalată de niciuna dintre celelalte metode, generând o adevărată ştiinţă legată de construcţia şi aplicarea sa, credem că centrul de greutate al acestei discipline a pihologiei aplicate, psihodiagnoza, trebuie să rămână testarea psihologică.

1.3. Marile domenii ale psihodiagnosticului

Prima arie în care s-a produs dezvoltarea şi utilizarea testelor mentale (pentru a respecta termenul propus de Cattell) a fost câmpul educaţional, şcoala, prin bateria creată şi publicată de Binet şi Simon în 1905. La peste un secol de la lansarea primei *Scale metrice a*

inteligenţei, şcoala rămâne în continuare marele beneficiar al psihodiagnozei. Utilizarea acesteia pe terenul educaţiei a avut şi are şi astăzi funcţii complexe:

- Depistarea copiilor cu o dotare intelectuală submedie (debilitate mintală, intelect de limită), sau cu alte caracteristici neuropsihiatrice, pentru a li se crea condiţii favorabile de tratament psihopedagogic diferenţiat, prin şcoli speciale sau prin integrarea lor în învăţământul de masă.
- Depistarea supradotaţilor, pentru instituirea unei pedagogii a excelenţei (curriculum diferenţiat).
- Diagnosticarea problemelor de învăţare şi a celor comportamentale (de conduită), pentru a stabili programe educaţionale adecvate.
- Selecţia educaţională prin teste standardizate la colegii sau universităţi (nu ca o procedură unică şi exclusivă, ci în asociere cu alte metode, cum ar fi scrisorile de recomandare, mediile din timpul anilor de studiu la anumite discipline şcolare, interesele elevului în afara şcolii etc.).
- Testele educaţionale (docimologice sau de cunoştinţe), pentru a verifica fie gradul de atingere a obiectivelor operaţionale, fie pentru a face un diagnostic – local, regional sau naţional – al învăţământului însuşi ca sistem, fie pentru a detecta cunoaşterea minimă prin posesia căreia se poate face o certificare a educaţiei primite într-un ciclu sau tip de şcoală.

Prin intermediul psihodiagnosticului şcolar se poate asigura personalizarea învăţământului şi tratarea diferenţiată a elevilor, determinarea nevoilor de educaţie, orientarea şcolară şi profesională, selecţia elevilor pentru programe speciale, determinarea potenţialului de învăţare şi promovarea metodelor de diagnostic formativ, controlul şi autocontrolul procesului educativ, surprinderea elementelor motivaţional-afective legate de învăţare, radiografierea grupului şi a dinamicii sale etc.

Credem că zona celei mai vaste şi mai fructuoase întâlniri dintre ştiinţa testelor (psihometria) şi ştiinţa examinării (docimologia, doxolgia sau docimastica) este cel al evaluării. Chiar *randamentul şcolar*, ca raport dintre aptitudinile, interesele elevului şi gradul de acoperire al obiectivelor operaţionale, exprimate prin calificative, note sau medii şcolare, este locul unde testarea psihologică şi cea educaţională se întâlnesc în modul cel mai fericit.

Domeniul sănătăţii, în general, şi cel al sănătăţii mintale în special, presupune o fericită sinteză între funcţia de psihometrician a psihologului clinician şi cea de diagnostician, unde contribuţia lui rămâne „esenţială şi fundametală" (Ionescu, op. cit., p. 55), cu cea de psihoterapeut şi de cercetător. În echipa medic-psiholg, cel din urmă ar avea o „situaţie privilegiată în cercetare, unde el nu face, asemenea medicului, un examen, nu acordă o consultaţie, ci întreprinde o investigaţie, el desfăşoară o activitate de explorare, ceea ce îi asigură un statut special în cercetere" (idem, p. 59). Locul psihologului nu este numai în clinica de boli nervoase, deoarece prezenţa sa în sănătate duce la o reconsiderare a actului medical în ansamblul său prin repunerea în drepturi a subiectivităţii bolnavului. De aceea asistăm la o extindere progresivă a ariei de competenţă a acestui tip de specialist, şi simultan a calităţii prestaţiei sale, motiv pentru care în multe dintre ţările vestice stagiile de formare pentru psihologul clinician sunt pe deplin comparabile cu cele ale medicului, pentru a nu mai aduce în discuţie importanţa acordată pregătirii psihologice a medicului însuşi.

Psihologul clinician intervine diagnostic şi terapeutic în:

- Aprecierea normalităţii/anormalităţii dezvoltării neuropsihice a copilului şi adolescentului (clinica pediatrică).
- În investigarea dezordinilor de personalitate, a stărilor reactiv-anxioase, a tentativelor de suicid sau a stărilor de criză.
- În comportamentul deviant, asocial, antisocial sau delictual.
- În diagnoza şi tratamentul tulburărilor alimentare (anorexie, obezitate, bulimie).
- În evaluarea dezordinilor neurologice sau psihiatrice, prin calculul indicelui de deteriorare, ca şi în recuperarea unor funcţii mintale deteriorate parţial sau total.
- În diagnoza şi recuperarea problemelor produse de involuţie, sau de patologia asociată îmbătrânirii.
- În expertizarea capacităţii de muncă sau a deficienţelor de intelect pentru acordarea certificatelor de persoană handicapată.

Domeniul sănătăţii este cel în care funcţia diagnostică şi cea prognostică sau recuperatorie se întâlnesc plenar, făcând din psiholog nu numai un om care aplică teste (de inteligenţă, memorie, personalitate, aptitudini, valori, interese etc.), ci un specialist care acumulează o experienţă preţioasă în prevenţie şi în terapie, oriunde intervine o disfuncţie, sau acolo unde ea încă nici nu există, pentru a face posibilă nu numai personalitatea optimală, ci şi pe cea maximală.

În toate aceste domenii s-a conturat o zonă de elecţie a muncii psihologului, cea de *consiliere*, a sfatului avizat, pentru o tot mai mare diversitate de situaţii. Cum ar putea fi acest copil mai puţin conflictual în grup? Ce carieră ar fi cea mai potrivită pentru persoana X? Ce activităţi i s-ar putea recomanda după pensionare lui Y, pentru a se potrivi cu structura sa cognitivă şi cu interesele sale? Cum ar putea depăşi solicitările prea stresante o personalitate de tip A, pentru a preveni un infarct? Dar diabeticul sau hemofilicul – ce şanse au să coabiteze cu aceste boli cronice? Sau ce risc există pentru un cuplu care are un copil cu o boală genetică (sindrom Down, de exemplu) de a avea un al doilea copil cu aceeaşi afecţiune (sfat genetic)?

Domeniul muncii oferă un câmp de acţiune imens psihodiagnosticianului, care trebuie:

- Să evalueze forţa de muncă, pentru a face o selecţie în domeniul profesiunilor.
- Să prevină fluctuaţia forţei de muncă şi să limiteze accidentele de muncă.
- Să avizeze periodic activităţile cu risc crescut (lucrul la înălţime, portul de armă etc.).
- Să ofere criterii valide de promovare şi să ajute la construirea carierei.
- Să identifice structura aptitudinală, cea a factorilor de personalitate şi a intereselor pentru a ajuta la reprofesionalizare, adică la reconversie profesională.
- Să asiste tehnic reciclarea şi reprofilarea, într-o lume în care schimbarea devine condiţia sa permanentă.
- Să contribuie la ameliorarea climatului de muncă prin cunoaşterea dinamicii grupurilor.
- Să se implice în adaptarea omului la maşină şi a maşinii la om (aspecte ce ţin de ergonomie, destinată să amelioreze mijloacele puse la îndemâna muncitorilor şi să le facă compatibile cu caracteristicile lor fiziologice şi psihologice).

În *lumea afacerilor* testarea şi evaluarea psihologică au o pondere în creştere accentuată, deoarece de ele beneficiază întreg domeniul resurselor umane: selecţia funcţionarilor şi a celorlalţi angajaţi se bazează pe teste de aptitudini, achiziţii, interese, motivaţii, valori, care ajută nu doar la angajare, ci şi la promovare, transfer, creşterea performanţelor, satisfacţia muncii, alegerea pentru stagii de formare aprofundată. Studiul comportamentului consumatorului, marketingul şi vânzarea produselor, managementul, reclama etc., toate intră în această sferă a lumii afacerilor, care are foarte multe suprapuneri cu cea a muncii, a educaţiei şi a sănătăţii. De fapt, toate domeniile activităţilor umane, de la cele deja amintite,

la transporturi, armată, poliţie şi până la creaţie şi artă beneficiază de câştigurile din zona diagnosticului şi a evaluării psihologice.

Nu există nici un domeniu al activităţilor umane, nici un fel de manifestare a umanului care să nu poată servi ca punct de plecare pentru diagnoza psihică. Remarcabil este faptul că în acest domeniu experienţa practică şi cercetarea ştiinţifică, ce duce la cristalizări teoretice, se presupun permanent. Iată cuvintele generoase ale promotoarei psihodiagnosticului în România, Ursula Şchiopu: „*Prin diagnoza psihică se pune naturii umane o întrebare, se caută un răspuns care devine într-un fel valabil pentru mai ample probleme esenţiale privind personalitatea umană. Actul diagnozei psihice este doar momentul prim al unei activităţi mai complexe, aceea de recuperare şi de recondiţionare psihologică a capacităţii umane şi a dimensiunilor funcţionale optime ale personalităţii umane"* (op. cit., p. 45). Aceasta înseamnă că un psihodiagnostician bun nu iese ca atare de pe băncile facultăţii, de unde el ia doar cadrul general al pregătirii sale. El se formează în timp, prin acumulare de experienţă practică diversă (diagnostică, formativă şi terapeutică), de cunoaştere teoretică pentru a putea pune ordine şi extrage informaţie relevantă din datele culese. Un bun psihodiagnostician are în spate ani de muncă, fişiere cu mii de cazuri care îi oferă baza de date pentru construcţia de bareme şi etaloane aduse la zi, are instrumente diverse şi verificate practic, cu care poate să rezolve o gamă largă de probleme concrete. Formarea sa prin zeci de ore de practică pe lângă un psiholog experimentat şi certificarea competenţei sale sunt alte aplicaţii importante ale psihodiagnozei.

1.4. Probleme etice ale psihodiagnozei

Discuţiile în legătură cu responsabilităţile sociale ale muncii psihologului au fost frecvent reluate la majoritatea congreselor internaţionale, fiind concretizate într-un cod deontologic al Asociaţiei Internaţionale a Psihologilor, la care şi România a aderat.

Trebuie să arătăm că perioada de dezvoltare explozivă a testelor standardizate în America dintre cele două războaie mondiale a avut mai multe consecinţe practice, printre care şi publicarea unei sinteze bibliografice, mereu reîmbogăţită, numită *Educational, Psychological and Personality Tests of 1933 and 1934,* care va fi de fapt prima ediţie a ceea ce mai târziu se va numi *Mental Measurements Yearbook* (Buros, 1938), ce lista miile de teste apărute, sortându-le tematic, dar şi după criterii de calitate şi performanţe tehnice.

Interesul pentru copii supradotaţi s-a revigorat odată cu lansarea în spaţiu de către fosta Uniune Sovietică a primului satelit Sputnik (4 octombrie 1957), când, la un an, Congresul american a alocat sume importante pentru crearea de teste de abilităţi şi de

aptitudini „în graba de a detecta studenţi dotaţi şi talentaţi academic” (Cohen, Swerdlik şi Smith, 1992, p. 53).

Extinderea programelor de testare pe scară largă, combinată cu utilizarea excesivă a testelor de aptitudini şi de personalitate în toate domeniile, au redeşteptat îngrijorarea publică legată de utilizarea şi eficienţa testelor psihologice. Primele *Recomandări tehnice pentru testele de achiziţii* au fost publicate de Asociaţia Psihologilor Americani (APA, datând din 1895) în anul 1954, precedate fiind de *Standardele etice ale psihologilor* (1953) şi urmate de *Standarde pentru testele educaţionale şi psihologice şi manualele lor* în 1966. Acestea vor fi urmate de *Principiile etice ale psihologilor* (1981) şi de o *Casetă cu principiile psihologilor* (1987) sau de un *Cod al practicilor de testare în educaţie* (1988). Recomandări speciale au fost făcute în legătură cu testarea membrilor minorităţilor culturale şi lingvistice, a persoanelor cu dizabilităţi, ca şi în legătură cu nou apăruta ramură a administrării, scorării şi interpretării computerizate a testelor.

Standardele din 1985 indică şi câteva dintre drepturile celor testaţi.

- Dreptul de a-şi da *consimţământul* asupra testării: cu ce instrumente sunt testaţi, cum vor fi utilizate datele şi cui vor fi remise informaţiile despre rezultatele obţinute.
- Dreptul de *a fi informat* asupra rezultatelor: pentru a contrabalansa tendinţa încetăţenită de a da celui testat cât mai puţine informaţii despre performanţa sa (cu scopul de a nu produce o creştere a anxietăţii sau de a nu genera o criză), s-a hotărât să se dea un feedback mai realist subiectului, centrat nu numai pe rezultatele pozitive obţinute de el, ci şi pe cele negative.
- Dreptul la *non-invazia vieţii intime*, private: există o informaţie privilegiată, protejată prin lege, cum ar fi cea obţinută din relaţia avocat-client, medic-pacient, preot-credincios, soţ-soţie şi psiholog-client.
- Dreptul la *cea mai puţin stigmatizantă etichetare*, care să nu lezeze demnitatea umană a clientului.
- Dreptul de *păstrare a confidenţialităţii* datelor: utilizatorul trebuie să-şi ia toate precauţiile în depozitarea acestora în spaţii încuiate, în dulapuri de oţel sau în computere parolate, care să limiteze accesul doar la persoanele autorizate.

O parte dintre aspectele prezentate mai sus se regăsesc stipulate şi de legislaţia altor ţări în care practica profesiei de psiholog a atins anumite standarde. Astfel, Societatea Franceză de Psihologie a creat un cod deontologic similar celui american, cu precizarea unor

reguli de conduită morală ce pot fi considerate de valoare universală. Redăm doar câteva dintre ele, pentru o posibilă analiză comparativă, inclusiv cu legislaţia românească.

- *Etica profesională* stipulează că „în exercitarea profesiunii, psihologului îi este interzis orice act sau cuvânt care ar putea aduce atingere demnităţii umane".
- *Secretul profesional*, care „trebuie păstrat în vorbe, în conversaţie sau în depozitarea documentelor în legătură cu tot ceea ce psihologul a văzut, a auzit sau a înţeles în cursul practicii sale."
- *Respectul altuia*: „psihologul nu trebuie să se folosească de mijloace profesionale pentru a-şi asigura avantaje personale".
- *Informarea ştiinţifică* este un principiu care stipulează obligaţia psihologului de a se informa continuu asupra evoluţiei cunoştinţelor din domeniul său şi „să aibă grijă de a comunica ştiinţa sa cât mai complet posibil, într-un spirit de exactitate şi de adevăr".
- *Autonomia tehnică* obligă psihologul să rezilieze orice angajament pentru care nu se simte apt prin tehnicile pe care le deţine, el având în întregime „responsabilitatea alegerii metodelor pe care le foloseşte".
- *Independenţa profesională*, care nu neagă munca de echipă a psihologului, ci prevede doar faptul că el „nu trebuie să accepte condiţii care ar putea să atingă independenţa sa profesională, adică l-ar împiedica să aplice principiile deontologiei profesionale".

Consecinţă a legii 213/2004 care reglementează profesia de psiholog în România, a fost elaborat şi un **Cod deontologic al profesiei de psiholog cu drept de liberă practică** care instituie regulile de conduită profesională ale practicianului sau cercetătorului din acest domeniu de activitate. Codul deontologic este destinat în mod expres orientării şi reglării doar a acelor activităţi în care psihologii sunt angajaţi profesional, şi nu se referă la viaţa particulară a acestora, care poate fi luată în discuţie doar atunci când aceasta ar aduce prejudicii profesiei de psiholog.

Codul deontologic se sprijină pe următoarele principii fundamentale:

- ***Respectarea drepturilor şi demnităţii oricărei persoane***: prin activitatea lor profesională psihologii respectă trăirile, experienţele, valorile, opţiunile celorlalţi. Ei nu prejudiciază imaginea publică a clienţilor lor şi vor utiliza un limbaj care indică respectul autentic al demnităţii umane. Utilizarea informaţiilor obţinute pe cale profesională nu se va face niciodată în detrimentul drepturilor fundamentale ale omului, fie el beneficiar de servicii psihologice, participanţi la cercetare, angajaţi sau studenţi.

- ***Principiul responsabilităţii profesionale şi sociale***: psihologii vor încuraja starea de bine a clienţilor lor şi vor respecta dreptul acestora de a sista, fără nici o justificare, participarea la serviciul furnizat sau la activităţile de cercetare, în calitate de subiecţi. Psihologii vor participa direct la dezvoltarea psihologiei ca ştiinţă, prin menţinerea celor mai înalte standarde în materie, dar şi la dezvoltarea societăţii în general. Ei vor respecta legile şi reglementările societăţii sau comunităţii în care trăiesc şi vor face tot posibilul pentru a fi impuse şi respectate standardele şi principiile etice.
- ***Principiul integrităţii profesionale***: potrivit acestui principiu psihologul nu va practica nici un fel de discriminare. El vor promova acurateţea, obiectivitatea şi onestitatea sau buna credinţă în raporturile cu clienţii lor. De asemenea, el va evita recompensele exagerate, conflictele de interese care i-ar reduce imparţialitatea sau care ar putea aduce atingeri profesiei sale sau imaginii celorlalţi psihologi. El va fi deschis şi va recunoaşter cu onestitate limitele propriilor competenţe.

În mod foarte explicit, codul deontologic analizează câteva categorii de **standarde etice generale**, cum ar fi cel al competenţei, standardele cu privire la relaţiile umane, standardele de confidenţialitate, standardele de conduită colegială, standardele de înregistrare, prelucrare şi păstrare a datelor, standardele relative la onorarii şi taxe, ca şi standardele pentru declaraţii publice şi reclamă. O categorie specială o constituie **standardele specifice**, care se referă în mod explicit la educaţie şi formare profesională, la evaluare şi diagnoză, la cercetarea ştiinţifică şi la valorificarea rezultatelor acesteia.

Astfel, pentru *evaluare şi diagnoză* sunt formulate prevederi exprese despre prezentarea caracteristicilor psihologice, condiţiile de utilizare a testelor psihologice, obţinerea consimţământului pentru evaluare/ diagnoză, construcţia instrumentelor, interpretarea rezultatelor, calificarea necesară pentru a dezvolta o testare/ evaluare, responsabilităţile legate de administrarea instrumentelor şi prezentarea rezultatelor pentru cei evaluaţi, ca şi respectarea dreptului de proprietate intelectuală. Codul deontologic furnizează indicaţii la fel de detaliate pentru *cercetarea ştiinţifică* şi valorificarea rezultatelor acesteia, în acord cu standardele internaţionale existente relative la obţinerea acordului de a dezvolta o cercetare, obţinerea consimţământului, utilizarea suporturilor de înregistrare audio-video, evitarea unor categorii de subiecţi, utilizarea animalelor în cercetare, plagiat, abuz de status, transmiterea şi protejarea datelor, onestitatea ştiinţifică şi regulile de bună conduită în cercetarea ştiinţifică.

CAPITOLUL 2

TESTUL PSIHOLOGIC, INSTRUMENTUL DE BAZĂ AL PSIHOLOGIEI APLICATE

2.1. Scurt istoric al apariţiei şi dezvoltării testelor

Nici o altă metodă a psihologiei aplicate nu a avut o contribuţie mai efectivă în impunerea ei în aproape toate domeniile de activitate umană, dar şi în declanşarea unor dispute care au culminat cu o puternică atitudine „antitest", iniţial în fosta U.R.S.S., apoi în Statele Unite ale Americii (ţara în care practica testelor a luat cea mai mare amploare), dar şi în Ţările de Jos, ulterior. Este de aceea întemeiată metafora Ursulei Şchiopu, care consideră testul „copilul scandalos al psihologiei, ce mai tulbură şi astăzi spiritele" (1976, p. 104).

Intuit încă de Francis Galton la sfârşitul secolului al XIX-lea, care a încercat fără succes să dezvolte un test pentru a determina abilităţile mintale prin măsurarea unor caracteristici perceptual-motrice, termenul a fost introdus în psihologie în 1890 de James McKeen Cattell. Consacrarea metodei testelor vine însă prin Alfred Binet, care foloseşte „*Scara Metrică a Inteligenţei*" (1905) pentru determinarea dezvoltării intelectuale a copiilor, extinderea utilizării lor datorându-i-se lui Munsterberg (determinarea aptitudinilor în vederea selecţiei profesionale).

La mai mult de un deceniu (1916) de la publicarea testului Binet, Louis Terman de la Universitatea Stanford a elaborat versiunea engleză a acestuia, cunoscută sub numele de „*Scala Stanford-Binet*". Chiar după această dată Henry Goddard a realizat propria sa traducere a acestei scale, deşi anterior (încă din 1912) el dezvoltase o metodă de determinare a defectivităţii mintale prin testarea imigranţilor şi selectarea acelora care urmau să fie deportaţi. Incluzând probe care cereau familiarizarea cu scrisul, dar şi cu cultura americană, Goddard a descoperit (1917) că peste 80% din imigranţi erau defectivi mintal, dând şi un clasament pe naţiuni… Era o primă utilizare greşită a testelor, dar nu şi ultima.

În timpul primului război mondial (1917, 1918), o echipă de psihologi, care l-a inclus şi pe Goddard, a dezvoltat două baterii de teste pentru selecţia primară şi repartizarea pe arme a recruţilor: *Army Alpha Test* şi *Army Beta Test*, pentru cei care aveau o bună, respectiv slabă posesie a limbii engleze. Aplicate în condiţii de stres, improprii după standardele de astăzi (camere aglomerate, instructaj greu audibil de către toţi subiecţii), rezultatele acestor teste l-au determinat pe Robert Jerkes să concluzioneze că vârsta mintală, pentru 47% dintre recruţi, nu depăşea 13 ani… Statistici „ştiinţifice", întemeiate pe astfel de date oferite de testări în

masă, au dus la apariţia *Legii Imigraţiei* din SUA anilor 1924, care stabilea „cote" de imigraţie pentru diferite naţiuni, după rezultatele obţinute la testele aplicate anterior altor imigranţi. Rezultatele la teste „au fost de asemenea utilizate de unii pentru a argumenta segregarea negrilor şi, în unele state, unii oameni, după scorul testelor de QI, au primit eticheta de 'imbecil' şi au putut fi sterilizaţi împotriva dorinţei lor, sau chiar fără ca ei să ştie" (Bernstein, Roy, Srull şi Wickens, 1988, p. 372).

Toate acestea nu au putut însă opri evoluţia testului psihologic, a cărui extindere progresivă a făcut ca rolul său să fie considerabil şi în creştere şi azi. Copil al altor metode, cum ar fi observaţia şi experimentul, „testul a însemnat un progres enorm în metodologia cunoaşterii structurii individuale şi prin aceasta a condiţionat extinderea aplicaţiilor psihologiei şi consolidarea pe bază experimentală a psihologiei diferenţiale" (Holban, 1973, p. 142). Şi aceasta deoarece testele au o valoare practică dovedită, generată mai ales de precizia, obiectivitatea şi relevanţa informaţiilor obţinute. Testele psihologice au un mare randament, oferă, în condiţii de costuri materiale şi de timp scăzute, o informaţie care ar fi fost greu de obţinut prin alte metode. Astăzi este tot mai evident că nu testele înseşi trebuie incriminate, deoarece la capătul unei lungi evoluţii au standarde foarte înalte relative la elaborare, utilizare şi interpretare, incomparabile cu oricare altă metodă, ci modul lor de utilizare. Ele sunt doar instrumentele perfecţionate ale psihologiei aplicate, dar maniera lor de folosire depinde în principal de utilizator, şi nu de caracteristicile lor intrinseci.

2.2. Definirea testelor

Termenul „test" este preluat din limba engleză, unde *to test* înseamnă „a pune la probă", „a încerca", chiar dacă, pe filiera latină, cuvântul exista şi în limba română (*testamentul* este un mod de probare a voinţei cuiva în legătură cu destinaţia bunurilor sale după moarte). Mai mult, termenul nu este specific psihologiei: medicina foloseşte tuberculina pentru a testa reacţia organismului la bacilul Koch, chimia are testul hârtiei de turnesol, fizica testul descompunerii luminii şi analiza spectrografică iar sociologia testează opinia publică prin chestionare. În sens larg, chiar o ipoteză sau o teorie ştiinţifică devin valide prin testarea lor în realitate (o eclipsă de soare *a testat* ipoteza curbării razei de lumină în preajma corpurilor cu mase mari, element central al teoriei lui Einstein), în timp ce Freud sau psihologia cognitivă vorbesc de capacitatea unei persoane de *a testa realitatea*.

Pentru a surprinde specificul testelor şi al testării psihologice, vom semnala câteva definiţii, aceasta deoarece nici în această privinţă nu s-a ajuns la un consens. Testul este deci:

- Probă, instrument pentru diagnosticarea de caracteristici psihice (Şchiopu, *Dicţionar enciclopedic de psihologie*, 1997).
- Probă utilizată mai ales în psihologia diferenţială, care permite descrierea comportamentului unui subiect într-o situaţie precis definită (instructajul testului), prin raportare la comportamentul unui grup de subiecţi plasaţi în aceeaşi situaţie (*Grand dictionnaire de psychologie*, 1994).
- Măsurătoare obiectivă a unui eşantion de comportament (Anastasi, 1976).
- Procedeu sistematizat de măsurare a comportamentului unor persoane şi de descriere a acestuia cu ajutorul unor scări numerice sau a unui sistem de categorii (Cronbach, 1966).
- Probă determinată, implicând o sarcină de executat, identică pentru toţi subiecţii examinaţi. Dispune de o tehnică precisă pentru aprecierea succesului şi pentru notarea numerică a reuşitei (Piéron, 1963).
- Situaţie standardizată, servind drept stimul unui comportament. Acest comportament este evaluat şi comparat statistic cu al altor persoane plasate în aceeaşi situaţie, ceea ce permite clasarea subiectului, fie cantitativ, fie tipologic (Pichot, 1984).
- Procedură standardizată prin care se formează un eşantion de comportament care va fi descris prin categorii sau scoruri. În plus, multe teste au norme sau standarde care fac posibilă utilizarea rezultatelor în predicţia altor comportamente, mai importante (Gregory, 1992).
- Instrument al metodei experimentale, organizat sub forma unor probe standardizate din punct de vedere al conţinutului, al condiţiilor de aplicare şi al modalităţii de apreciere a rezultatelor, instrument care este folosit în stabilirea unei anumite variabile (Holban, 1973).

Vom reţine din această multitudine de definiţii elementele sintetice cele mai semnificative, esenţiale pentru înţelegerea specificului unui test psihologic care:

- este probă, procedură, procedeu sau instrument standardizat;
- care serveşte drept stimul pentru producerea şi măsurarea unui eşantion de comportament;
- care va fi apreciat prin raportarea şi compararea cu o populaţie de referinţă, testată în condiţii identice, în raport cu care s-a construit sistemul de apreciere (barem, etalon, norme, categorii tipologice);

- funcţia lui fiind în acelaşi timp diagnostică (apreciază starea de fapt), dar şi prognostică (anticipează categorii viitoare de comportamente mai importante decât cele măsurate efectiv);
- pentru îndeplinirea acestor funcţii ele trebuind să aibă calităţi psihometrice specificate (fidelitate, validitate, sensibilitate sau forţă de discriminare).

2.2.1. Eşantionul de comportament

Ca şi biochimistul, care ia un eşantion de sânge sau mostre de apă din locuri diferite pentru a determina, prin analize, starea de sănătate sau potabilitatea apei, psihologul alege pentru testare eşantioane mici, dar cu grijă selecţionate, din comportamentul unei persoane. Aceasta deoarece el trebuie nu numai să diagnosticheze, ci mai ales să facă predicţii valide. Cum investigarea comportamentului în totalitatea sa este imposibilă, cu atât mai mult cu cât el se desfăşoară pe scara timpului, având doar o constanţă relativă, psihometricianul trebuie să aleagă eşantioane de comportament mai mici, dar strâns şi semnificativ legate de comportamentul marii mase studiate. Această unitate ia în test forma *itemului,* ce este „constituit dintr-o situaţie stimul care solicită un răspuns ce poate fi evaluat separat de restul testului" (Albu, 1998).

Nu este obligatoriu să existe o corespondenţă sau o similaritate foarte mare între comportamentul prezis şi itemii testului: în timp ce corespondenţa dintre proba de examen la conducerea auto şi conducerea efectivă este deplină, un test proiectiv poate prezice, din mărimea, forma, culoarea sau mişcarea invocate în răspunsurile la test, importante caracteristici ale personalităţii celui examinat. „Testul nu este nici analitic, nici sintetic, ci analog cu situaţia reală, în sensul că modelează conţinutul psihologic al activităţii reale, dar nu şi forma exterioară a acestuia", afirmă Havârneanu (op. cit., p. 90). Aceasta ne îndreptăţeşte să afirmăm că dincolo de deosebirile de formă, trebuie să existe o identitate internă sau de conţinut între test şi sarcina reală în raport cu care se face predicţia, de unde rezultă şi valoarea constructivă a testului.

Aşa cum apreciază Anne Anastasi, de multe ori predicţia se bazează pe diagnosticul unui potenţial sau capacităţi: „nici un test psihologic nu poate face mai mult decât să măsoare un comportament. Dacă un asemenea comportament poate servi ca un indicator efectiv al altui comportament, acesta poate fi determinat numai prin cercetare empirică" (1976, p. 25).

2.2.2. Standardizarea

Standardizarea este o condiţie fundamentală a testului psihologic, unde orice variaţie a condiţiilor produce o variaţie a rezultatelor. Standardizarea se referă atât la *conţinut* (proba este identică cu sine însăşi în ce priveşte nu numai conţinutul itemilor, dar chiar şi forma lor de prezentare şi foaia de răspuns), ci şi la *condiţiile de aplicare* (asemănarea până la identitate dintre condiţiile în care s-a experimentat şi cele în care se aplică proba) şi în *modul de corectare, cotare şi apreciere a probei* de către autorul, respectiv utilizatorul unui test. Fiind o metodă de bază a psihologiei diferenţiale, testul poate fi considerat variabila independentă prin care o însuşire psihică (variabila dependentă) este observată prin examinarea persoanelor care sunt testate (păstrarea constanţei situaţiei, elementul de variabilitate fiind persoanele).

Deoarece în psihologie pluri-determinarea face dificilă izolarea fenomenelor, cerinţa *uniformităţii şi* a *identităţii de procedură* este mai stringentă ca în alte domenii. Pentru a asigura uniformitatea, autorul testului dă indicaţii precise şi detaliate despre cum se administrează şi cum se scorează rezultatele testului, indicaţii ce vor fi respectate foarte strict, pentru că altminteri rezultatele culese nu vor mai fi compatibile cu cele obţinute pe eşantionul standard (de etalonare sau eşantion normativ). Materialele folosite, modul lor de prezentare, instructajul subiecţilor, limitele de timp, demonstraţia prealabilă, modul de manevrare a materialelor, toate sunt specificate şi respectate întocmai. Uniformitatea se referă şi la rata vorbirii examinatorului, la tonul, inflexiunile vocii sau expresia sa facială.

Deoarece un test nu are indicatori predeterminaţi pentru aprecierea succesului sau a eşecului, un alt aspect al standardizării se referă la *norme*, adică la stabilirea valorilor medii (normale) ale performanţelor, în funcţie de care, plecând de la scorurile brute (numărul de itemi rezolvaţi corect, timpul, viteza, numărul de erori etc.) se stabilesc scorurile standard. Aceasta presupune activitatea de etalonare, care se face pe loturi mari, reprezentative statistic, cărora li se determină media, mediana, abaterea standard, frecvenţele brute şi cumulate, în funcţie de care se stabilesc cotele standard. Performanţa oricărui individ nou examinat va fi raportată la acest cadru de referinţă. Pentru ca acest lucru să fie posibil este obligatorie uniformitatea şi identitatea procedurilor de aplicare a testului în noul caz, cu situaţia de etalonare originară.

2.2.3. Fidelitatea

Dacă cineva şi-ar măsura greutatea, citind la un cântar 80 de kg., iar la altul 92 de kg. în aceeaşi zi, nu ar şti nici care este greutatea sa reală, nici care dintre cântare este bun (dacă este vreunul). Cu atât mai mult ne vom îndoi de indicaţiile cântarului respectiv cu cât, la două

măsurători succesive ale aceleiaşi persoane, el va indica valori diferite. Un test, ca oricare altă scală de măsurare, trebuie să fie fidel, adică egal cu sine însuşi, rezultatele unor măsurători repetate trebuind să fie cel puţin stabile, dacă nu identice.

Cu cât fidelitatea unui test este mai mare, cu atât mai mare siguranţă avem că rezultatele lui sunt reale, nefiind în mod semnificativ afectate de schimbări aleatoare, cum ar fi contextul, situaţia de testare, starea subiectului, particularităţile examinatorului etc. „Utilizat în psihometrie, termenul de fidelitate (*reliability* în engleză) întotdeauna înseamnă consistenţă", afirmă Anastasi (1972, p. 27).

Există modalităţi speciale de determinare a fidelităţii. Acestea sunt:

- Procedura *test – retest* (corelaţia dintre rezultatele aplicării aceluiaşi test, în aceleaşi condiţii, aceloraşi subiecţi, eventual de către acelaşi examinator, în două momente de timp diferite, la distanţă relativ scurtă, pentru a nu apărea modificări în structura parametrului investigat).
- Calculul corelaţiei dintre rezultatele unui test şi *forma sa paralelă*, alternativă (în cazul în care ea există), aplicată în condiţii identice.
- *Tehnica înjumătăţirii* (*split-half*), prin care jumătatea pară a unui test (itemii cu număr par) este pusă în corelaţie cu jumătatea lui impară. Există şi alte metode de înjumătăţire.
- Corelaţia între modul în care psihologi diferiţi scorează acelaşi test, aplicat aceloraşi persoane, în aceleaşi condiţii (*fidelitatea interscoreri*).

2.2.4. Validitatea

Un test poate să fie fidel fără a fi însă şi valid, însuşire care se referă la faptul că el măsoară efectiv ceea ce şi-a propus. Nici un test nu are validitate scăzută sau ridicată în mod absolut, deoarece aceasta depinde de maniera în care el este utilizat. Galton a încercat să determine într-o manieră corectă acuitatea perceptivă sau sensibilitatea la durere, dar acestea s-au dovedit a nu fi determinări valide ale inteligenţei, aşa cum el a presupus.

Spre deosebire de fidelitate, care se poate determina relativ uşor, validitatea are nevoie de acumulări graduale de informaţie, din foarte diferite tipuri de investigaţii. Deoarece este un proces progresiv, dependent de timp, este mai potrivit să spunem că *un test se validează*, validitatea fiind acea determinare care permite să spunem în ce măsură el îşi îndeplineşte funcţiile pentru care a fost proiectat, în raport cu diversele sale domenii de aplicaţie. Prin validitate, testul se testează pe sine, făcând apel la diverse criterii externe, pentru a determina

dacă predicţiile sale au fost valabile şi în ce măsură, în funcţie de care vom şti ce încredere să îi acordăm.

Efectiv, validitatea (ca şi fidelitatea), este o corelaţie care se poate stabili în trei feluri:

1. Măsura în care un test acoperă un domeniu, adică *validitatea de conţinut*. Un test de cunoştinţe poate acoperi un mic sector al matematicii, sau unul larg, caz în care el este mai valid în raport cu acest domeniu.
2. Măsura în care scorurile la un test sunt în acord cu teoria care susţine trăsătura sau constructul măsurat, adică *validitatea de construct.*
3. Măsura în care testul corelează cu un alt test independent, recunoscut ca un bun instrument de măsură al aceluiaşi domeniu (inteligenţă, memorie, personalitate etc.), ceea ce reprezintă *validitatea relativă la criteriu*. Când scopul testului este predicţia unui anumit comportament, criteriul va fi măsurătoarea viitoarei performanţe şi corelarea ei cu testul iniţial, în acest caz vorbind de *validitatea predictivă*. Aceasta este o formă fundamentală de validitate, pentru că de cele mai multe ori psihodiagnoza are în vedere prognoza, adică anticiparea rezultatelor. Unii detractori ai metodei testelor susţin că cea mai bună predicţie a rezultatelor unei munci sau a unei activităţi ar fi rezultatele obţinute într-o perioadă de timp în chiar munca sau activitatea respectivă. Aceasta ar face inutilă chiar activitatea de selecţie, dar această idee nu poate fi utilizată practic din cauza marii cheltuieli de timp şi bani presupuse de punerea ei în aplicare.

La aceste trei caracteristici: standardizarea, fidelitatea şi validitatea testelor, se adăugă şi *sensibilitatea* sau *forţa discriminativă*, cât şi ideea de standardizare, etalonare şi normă, ce vor fi tratate mai detaliat în capitolele următoare, întrucât de ele depind înţelegerea şi buna utilizare a acestor puternice instrumente psihodiagnostice, care sunt testele psihologice.

2.3. Administrarea testelor

Necesitatea de a avea stabilite proceduri detaliate sau doar orientative pentru administrarea şi scorarea testelor psihologice şi educaţionale a fost recunoscută de toate organizaţiile care se ocupă cu testele. *American Educational Research Association* a stabilit 180 de standarde care accentuează asupra importanţei uniformităţii în administrarea şi scorarea testelor. Procedurile de testare depind fie de tipurile de teste (individuale sau de grup, cu şi fără limită de timp, cognitive, afective sau de personalitate, psihologice sau educaţionale etc.), fie de subiecţii supuşi examinării (vârstă, sex, cultură, grad de motivaţie, anxietate etc.).

Formarea specialiştilor care vor administra teste este una de durată, se face sub supravegherea unui psiholog cu experienţă şi presupune calităţi personale dublate de cunoştinţe detaliate despre instrumentele psihometrice utilizate. Cum variabilele situaţionale au o pondere în rezultatul final, ele trebuie controlate cât mai complet posibil.

2.3.1. Îndatoririle examinatorului înainte de administrarea testelor

Pentru a preîntâmpina efectul de surpriză (neindicat), testele psihologice vor fi programate şi anunţate din timp, pentru a evita plasarea lor după masă, în timpul programului de joacă sau în concurenţă cu alte activităţi mai interesante.

În acord cu particularităţile de vârstă, dacă şedinţa de testare ar putea depăşi o jumătate de oră la grădiniţă, o oră în primul ciclu şcoalar şi o oră şi jumătate în al doilea ciclu şcolar, atunci se vor organiza mai multe şedinţe. În cazul testelor educaţionale elevii vor fi anunţaţi din timp, pentru a se putea pregăti fizic (odihnă), emoţional şi intelectual. Acolo unde este cazul, se va cere consimţământul scris, care este un acord de principiu al examinatului sau al reprezentantului legal al acestuia că el consimte să fie examinat. În ţările avansate această problemă este reglementată prin lege.

Tot ca o fază preparatoare este şi *pregătirea prealabilă a examinatorului*, care trebuie să fi parcurs el însuşi testul o dată, apoi să memoreze la nuanţă şi în detaliu instrucţiunile, pentru a nu avea nesiguranţă şi ezitări în aplicarea acestuia. Pentru a crea o atmosferă informală, calmă şi destinsă (mai ales în examinările individuale) el trebuie *să pregătească şi să verifice cu minuţiozitate materialele necesare,* reducând la minimum căutarea şi organizarea lor în timpul şedinţei de testare. Dacă va folosi aparate, verificarea bunei funcţionări şi calibrarea lor periodică sunt obligatorii. În examinările colective (de grup), fiecare caiet de test, foaie de răspuns, creioane sau instrumente necesare trebuie atent verificate. Familiarizarea cu procedurile de examinare presupune practica asistată de specialistul calificat, care poate merge de la simpla observaţie, până la peste un an de instruire supervizată.

Asigurarea condiţiilor satisfăcătoare de testare obligă psihologul să verifice condiţiile de ordin fizic (aşezare, iluminare, ventilare, temperatură, nivelul zgomotului) şi psihice, prin care examenul să se desfăşoare confortabil. Anunţul: „*Se testează! Vă rugăm, nu deranjaţi!*” va descuraja pe cei din exterior să inoportuneze, deşi prezenţa unui supraveghetor la uşă, şi chiar încuierea ei pe perioada sesiunii de testare, ar fi şi mai de dorit.

Dacă în cazul examenelor individuale (mai costisitoare, dar şi mai edificatoare) se pot controla mai uşor variabilele externe, examinatorul alegând o cameră liniştită, potrivită

scopului şi lucrând singur cu subiectul (chiar dacă uneori prezenţa părintelui, a reprezentantului legal sau a poliţistului se impune), în examinările colective precauţiile vor fi mult mai mari, în special datorită larg răspânditei tendinţe de a trişa sau de a fi incorect, care poate vicia semnificativ rezultatele. Pentru aceasta există mai multe măsuri de siguranţă ce pot fi luate: distribuirea de teste cu acelaşi conţinut, dar aşezate în forme diferite, prezenţa unui număr suficient de supraveghetori atent instruiţi, care vor descuraja activ sau prin simpla lor prezenţă tendinţa spre fraudă. Aceştia sunt cu atât mai necesari cu cât grupul este mai mare, iar importanţa testării (miza ei) este mai ridicată, fiind de mare ajutor în distribuirea şi colectarea materialelor, în secretizarea foilor de răspuns (când se impune) şi la supraveghere pentru preveniţia oricărei situaţii nedorite.

Detalii aparent nesemnificative cum ar fi forma pupitrului, tipul de foi de răspuns utilizate (indicate a se folosi pentru subiecţi de peste 11 ani, pentru că ele încetinesc considerabil viteza de parcurgere a testului), prezenţa sau nu a încurajărilor sau a comentariilor examinatorului, pot afecta performanţa la un test şi de aceea standardizarea procedurii de aplicare a acestuia trebuie să meargă până la cele mai mici nuanţe. Orice modificare a acesteia trebuie notată şi luată în calcul la interpretarea rezultatelor.

2.3.2. Faza de administrare a testelor

Examenul şi testarea efectivă pun probleme specifice, una dintre cele mai importante fiind *stabilirea relaţiei (raportului) examinator – subiect (subiecţi).* Primul va trebui să declanşeze interesul, curiozitatea şi cooperarea explicită a celui examinat. Testele de abilităţi trebuie să aducă subiectul în situaţia de a da cea mai bună performanţă posibilă a sa, pe când la testele de personalitate, la chestionarele de opinii, atitudini şi valori, obţinerea sincerităţii răspunsurilor este fundamentală. La testele proiective, evocarea liberă, fără cenzură, este factorul cel mai important. Problemele motivării, constanţa în lucru, evitarea negativismului şi a distragerii atenţiei presupun tehnici de stabilire a contactului bine exersate. Copiii (cu precădere cei mici) şi bătrânii, persoanele vulnerabile sau cele foarte ruşinoase, timide, din medii defavorizate sau aparţinând altor culturi ori etnii, creează probleme specifice. La acestea se mai adaugă delincvenţii, recidiviştii, deţinuţii, bolnavii psihic (în special psihopaţii, dar şi anxioşii, emotivii, depresivii, schizoizii şi persoanele paranoiace), care pot dezvolta suspiciune, indiferenţă, atitudini marcate de cinism, agresivitate sau nesinceritate.

Dacă la vârsta şcolară mică prezentarea testului ca un joc poate fi soluţia optimă, la cea şcolară mare pârghiile cele mai indicate sunt prezentarea testării ca pe o competiţie cu sine şi cu ceilalţi iar la liceeni stimularea nevoii de stimă şi de prestigiu. De asemenea,

familiarizarea prealabilă cu sarcini asemănătoare celor din şedinţa de testare ar putea fi folosită ca mijloc de reducere a tensiunii anxioase. Examenul adulţilor pentru ocuparea unor posturi, selecţia în anumite profesii sau pentru cursuri de calificare reclamă mult tact, prudenţă şi experienţă din partea psihologului.

În timpul desfăşurării examenului psihologic, examinatorul va trebui să aibă multă grijă în a urmări cu stricteţe procedurile standardizate de aplicare a testelor, neavând voie să dea nici un fel de indicaţii, lămuriri suplimentare sau ajutoare, altele decât cele stipulate în mod expres de manualul testului. Excepţie fac şedinţele pentru determinarea potenţialului de învăţare (strategie formativă definită şi de Feuerstein în 1987), când se dau ajutoare în plus faţă de cele prevăzute, dar şi acestea într-o formă standardizată, cuantificabilă, pentru ca prin calculul diferenţei fazei pre-test şi post-test să se determine în mod obiectiv acest potenţial de învăţare.

Examinatorul trebuie să rămână în alertă, pentru a rezolva orice problemă urgentă care apare, să fie flexibil în atitudine, cu sensibilitate şi răbdare mai ales pentru subiecţii care au probleme (handicapaţi, hiperkinetici, copii mici, timizi etc.). Aiken (1997) face câteva recomandări valabile pentru aceste categorii speciale:

- psihologul să asigure timp suficient celui examinat pentru a înţelege şi răspunde;
- să permită suficient antrenament la itemii simpli, pregătitori;
- să folosească perioade mai scurte de testare;
- să observe oboseala şi anxietatea şi să le ia în calcul;
- să sesizeze şi să noteze deficienţele perceptiv-motrice (defectele de auz, văz, motrice, lateralizarea inversă);
- să folosească generos încurajarea şi întărirea pozitivă;
- să nu forţeze subiectul să răspundă când el nu mai doreşte.

În timpul testării, mai ales la testele cu răspunsuri la alegere din mai multe posibilităţi sau variante, este posibil ca o parte dintre răspunsurilşe bune să poată fi efectiv ghicite, sau determinate prin ceea ce Aiken numeşte „deşteptăciunea în test” (*test wiseness*). Prin aceasta se elimină opţiunile nepotrivite, prin semne adiţionale scăpate în formularea itemilor, ce permit judecăţi comparative între opţiuni. Există tehnici speciale de ameliorare a performanţelor la anumite teste, nu prin pregătirea specifică în domeniu, ci prin respectarea unor reguli (Aiken, op. cit., p. 57).

2.3.3. Faza post-test

După terminarea şedinţei de testare, examinatorul colectează şi pune în siguranţă materialele de testare şi foile de răspuns. Dă mici bonusuri, recompense copiilor sau celor mai anxioşi sau vulnerabili. Uneori informează părinţii sau persoana în cauză despre utilitatea ce va fi dată rezultatelor testului. De asemenea el promite să furnizeze informaţiile rezultate persoanei sau agenţiei care l-a angajat. În caz de examen colectiv, psihologul şi ajutoarele sale colectează toate materialele utilizate în examen, se asigură că nimic nu lipseşte şi abia după aceea permite retragerea subiecţilor, după care el va pune în ordine foile de răspuns şi se va pregăti să le scoreze.

2.4. Examinatorul şi variabilele situaţionale

Există mulţi stimuli externi ambigui care trebuie luaţi în considerare sau controlaţi de către examinator. Copiii preşcolari pun probleme speciale examinatorului, ca şi bătrânii, emotivii, ruşinoşii, anxioşii sau vulnerabilii.

Vârsta, sexul, rasa, statusul social, economic şi cultural, experienţa, antrenamentul, aparenţa exterioară şi altele sunt de luat de asemenea în calcul în ceea ce priveşte diferenţele pe care examinatorul însuşi le poate introduce. Uneori chiar nivelul lui de expectaţie poate genera profeţia care se autoîmplineşte sau *efectul Rosenthal*.

Situaţia de examinare are şi alte faţete. Anne Anastasi arată că recruţii abia încorporaţi dau rezultate mai slabe dacă sunt examinaţi imediat după încorporare, fără să fi avut timp să se acomodeze cu noul mediu. Uneori activităţile imediat precedente îşi pun amprenta asupra performanţei la test. Feedback-ul pozitiv sau negativ dat de examinator pe parcursul desfăşurării testului poate ameliora sau diminua rezultatele finale. Chiar dacă ponderea acestor factori nu este foarte mare, într-un program de testare bine conceput examinatorul trebuie să fie foarte atent la ei pentru a le minimiza influenţa. Probleme suplimentare apar şi atunci când o şedinţă de testare este precedată de un antrenament practicat cu teste similare.

Deoarece practica testelor a luat o amploare extrem de mare în toată lumea şi deoarece multe decizii importante se bazează pe rezultatele obţinute la teste, edituri specializate sau firme îşi oferă „serviciile” în antrenarea candidaţilor. Într-o manieră similară cu a iluzioniştilor care fac publice secretele confraţilor, sau cu a informaticienilor care introduc viruşi în programele calculatoarelor, psihologii au căzut în propria cursă, ceea ce îi obligă la precauţii suplimentare ce fac din psihodiagnoză un domeniu tot mai tehnicizat şi mai sofisticat.

CAPITOLUL 3

DIFICULTATEA ŞI SENSIBILITATEA UNUI TEST PSIHOLOGIC SAU EDUCAŢIONAL

3.1. Dificultatea unui test

Aplicarea unui test psihologic sau educaţional ridică o problemă tehnică foarte importantă şi anume dacă el este adecvat, ca şi grad de dificultate, populaţiei sau grupului respectiv. Dacă este prea uşor, majoritatea scorurilor la test vor fi mari şi curba rezultată, în formă de „j", va fi asimetrică spre dreapta; dacă va fi prea greu, curba în formă de „i" va fi asimetrică spre extrema stângă. Se apreciază că testul este adecvat unei populaţii doar în cazul în care repartiţia scorurilor sale este simetrică, generând o curbă de tip gaussian.

Problema dificultăţii unui test se pune şi în funcţie de tipul acestuia. Dacă el este un test de aptitudini, capacităţi, deprinderi sau cunoştinţe, această problemă este mult mai relevantă decât pentru testele de personalitate sau chestionarele de opinii, atitudini, interese, valori, caz în care termenul de dificultate ar putea fi înlocuit cu cel de accesibilitate.

Dificultatea testului depinde simultan de conţinutul său, ca şi de particularităţile subiecţilor investigaţi, ceea ce face ca în construirea unui asemenea instrument să existe o *fază iniţială* (definirea testului, crearea băncii de itemi şi construirea unei variante preliminare a acestuia), o *fază intermediară* (administrarea testului, analiza rezultatelor obţinute şi, prin analiza de itemi, depistarea erorilor, corectarea, reelaborarea, selectarea şi ordonarea itemilor) şi o *fază finală* (crearea normelor, a etaloanelor şi elaborarea manualului testului, vezi Albu, 1998, p. 199).

Dacă testele de aptitudini şi de cunoştinţe aplicate adulţilor iau în considerare un proces încheiat, variabilitatea fiind dată mai ales de studii, profesie sau sex, cele mai multe teste de inteligenţă, memorie, aptitudini speciale şi cunoştinţe, aplicate copiilor sau adolescenţilor, au în factorul vârstă cea mai importantă sursă de variabilitate. În acest sens, elementele de dificultate ale testului trebuie să evolueze paralel cu vârsta (să aibă *indice genetic*, după expresia lui Claparède), în cadrul fiecărei vârste el trebuind să aibă o cât mai bună acoperire a spectrului domeniului investigat. Aceasta trimite la puterea de discriminare a testului, adică la sensibilitatea lui.

Un test este uşor sau greu în situaţia când unităţile lui, fiind foarte puternic corelate între ele (test omogen), creează posibilitatea ca majoritatea itemilor să fie rezolvaţi ori corect, ori greşit, ceea ce va da curbe asimetrice spre dreapta sau spre stânga. Există şi posibilitatea

ca populaţia să cuprindă două grupuri distincte, care vor da scoruri polarizate după o curbă bimodală (cu două „cocoaşe”), ca în cazul testelor dihotomice: masculinitate-feminitate, extraversiune-introversiune, stabilitate-instabilitate, admis-respins. Multe teste de selecţie profesională nu urmăresc ierarhizarea de fineţe a persoanelor, ci „ruperea grupului” în categorii extreme: apt-inapt, admis-respins, capabil-incapabil. Aici testele folosite vor fi mai mult centrate pe dificultate (mică, medie, mare) şi mai puţin pe sensibilitate sau putere de discriminare. Aplicate la începutul sau la sfârşitul unui proces de formare, acelaşi test poate poate avea o curbă iniţială în formă de „i” (când deprinderea încă nu există), în formă de „j” (când în final aceasta s-a cristalizat pentru majoritatea subiecţilor), între care se interpune o repartiţie normală, simetrică, pentru perioada intermediară (vezi Albu, op. cit., p.185).

3.2. Sensibilitatea unui test

Asemenea lentilelor, a căror calitate depinde în principal de puterea lor de separaţie (numărul de linii independente discriminate pe o suprafaţă de un centimetru pătrat), sau a balanţelor de fineţe, ca cele analitice, capabile să dea subdiviziuni din ce în ce mai fine şi continue până la a cincea zecimală, caz în care se spune că sunt foarte sensibile, testele psihologice au această caracteristică în funcţie de numărul claselor pe care le pot delimita în interiorul unui grup. Cu cât numărul valorilor de scor este mai mare, cu atât ele discriminează (diferenţiază) mai bine între indivizi.

În afara acestei accepţiuni, Kline (1993) leagă sensibilitatea unui test de capacitatea sa de a produce scoruri diferite pentru subiecţii care diferă între ei în ce priveşte caracteristica măsurată. Deoarece sensibilitatea conjugă caracteristicile testului cu cele ale populaţiei măsurate, vom remarca faptul că testele mai lungi, cu itemi gradaţi ca dificultate (deci neomogeni) sunt mai sensibile, pentru că ele produc mai multe clase, şi deci mai multe valori ale scorurilor. Prin contrast, testele scurte sau cu itemi foarte omogeni, generează mai degrabă categorii de scor dihotomice, foarte contrastante, deci ele sunt puţin discriminative. Depinzând de forma repartiţiei pe care o dă scorurilor sale, testul cu curbe atipice (i, j sau bimodală) sunt puţin sensibile în zonele de mare aglomerare a scorurilor şi sensibile în rest: curba lui Gauss este tot mai discriminativă spre extreme, şi tot mai puţin sensibilă pe porţiunea sa centrală, unde aglomerarea rezultatelor dă un număr mai mic de scoruri.

Există o legătură foarte puternică între dificultatea unui test şi sensibilitatea sa. La fel cum în fotografie filmele foarte sensibile (care se impresionează la cantităţi mici de lumină) plătesc un preţ în ceea ce priveşte calitatea imaginilor (mai puţin nete, deoarece imaginea este tradusă pe peliculă prin grupuri mai mari de puncte), sau la fel ca în metrologie, unde balanţa

analitică îşi dovedeşte utilitatea pentru greutăţile mici, iar nu pentru kilograme sau tone, testele psihologice trebuie să stabilească un raport optim între dificultate şi sensibilitate.

Sensibilitatea presupune, pe de o parte, o cât mai bună acoperire a domeniului de conţinut al variabilei, pentru care testul trebuie să aibă grade diferite ale dificultăţii, pe de altă parte discriminarea de fineţe la nivelul fiecărui palier de dificultate considerat. Pentru a exemplifica, testele de inteligenţă WISC sunt operaţionale de la 5-6 ani la 14-16 ani, pentru fiecare an dându-se etaloane din 4 în 4 luni. Dar, în timp ce anumite subteste ale bateriei (*Informaţii, Comprehensiune*) dau un larg evantai de scoruri, încât etaloanele nu au goluri, altele (*Cifrele, Labirinturile*), din cauza numărului mic de itemi care le compun, discriminează foarte grosier pe anumite porţiuni ale scalei, la anumite vârste. Subtestul *Cuburilor*, o foarte validă expresie a inteligenţei concret-spaţiale, compus din 10 patternuri ce pot da (în forma iniţială) 55 de puncte de scor, are un prag înalt de intrare în probă la vârstele mici (5-6 ani), unde nu distinge fin coeficienţii de inteligenţă mai mici de 80, şi un prag jos la vârste mari (peste 13 ani), unde nu mai distinge supradotaţii (QI > 125/130). Pentru bateria WISC, combinarea optimă a celor două caracteristici (dificultatea cu sensibilitatea) se face pentru vârstele de 11-13 ani, unde ea are cea mai mare forţă discriminativă şi deci cele mai bune calităţi psihometrice (Zimmerman şi Woo-Sam, 1973). În mod similar, testul PM 38 (*Matricile Progresive Standard* ale lui Raven) are un prag înalt de intrare în probă la 5-7 ani şi unul jos după 14 ani, deci nu distinge bine insuficienţa ori supradotarea mintală, la vârstele mici, respectiv mari.

Sensibilitatea testelor depinde de mulţi factori (care pot fi identificaţi şi prin analiza de itemi), dintre care amintim:

- Numărul itemilor, care, aşa cum am arătat, creează posibilitatea stabilirii numărului de clase de scor, progresiv cu creşterea acestuia. Numărul maxim de clase este chiar numărul de itemi, atunci când ei nu sunt echivalenţi sau intercorelaţi. Dacă luăm exemplul testelor de inteligenţă, acestea ar trebui să măsoare QI-uri de la 40 la 160, având o lungime a amplitudinii scorurilor (cea care trebuie să acopere întreg spectrul de variaţie) de 120 de unităţi. Dacă subtestul are 40 de itemi, înseamnă că cea mai fină rezoluţie este de 3 unităţi de QI pentru un item; dacă am avea, ca la subtestul *Cifre*, 8 itemi pentru prezentarea directă şi 7 pentru prezentarea inversă, înseamnă că 120/(8+7) = 8, adică fiecare item acoperă 8 unităţi de QI, ceea ce înseamnă o discriminare mult mai grosieră, în care orice eroare de aplicare ar putea avea consecinţe însemnate asupra scorului la acest subtest. Pe de altă parte, este aproape imposibil a genera un test cu maximă sensibilitate (un punct de

QI să fie dat de câte un item separat), nu numai din cauza extensiei amplitudinii împrăştierii la 120 de unităţi, dar şi din cauza numărului de vârste distincte pe care testul le acoperă (de la 5 la 16 ani sunt 12 vârste). Aceasta ar da subteste neobişnuit de lungi şi de greu aplicabile (nemaniabile). De aceea, teste de inteligenţă cu o rezoluţie de până la 5 unităţi de QI pe item sunt încă acceptabile, deşi cele mai bune valori ar fi între 2 şi 3.

- Maniera de scorare a probelor este iarăşi o modalitate importantă de creştere a sensibilităţii testelor. La testele *Domino* (*D 48* şi *D 70*), pentru fiecare problemă (alcătuită din desenele unor piese de domino ce solicită stabilirea a două valori) se acordă doar un punct, performanţa maximă fiind de 44 de puncte. Imaginând un alt sistem de scorare, prin care să nu se „piardă" puncte, şi acordând un punct dacă un element din căsuţa dublă este corect, încă unul pentru al doilea, plus un bonus de un punct pentru simultaneitatea lor şi orientarea corectă pe verticală (neinversare), se obţin 44·3 = 132 de puncte de scor, cu care se baleiază un domeniu mai extins al variabilei, dar şi vârstele succesive. Similar, la subtestul *Cuburilor* din bateria WISC, se dau 4 puncte pentru patternul rezolvat corect şi încă maximum 3 puncte drept bonificaţie de timp, rezultând maximum 7 puncte per item şi 55 de puncte în total. Deoarece rezolvările parţial corecte nu se iau în considerare, rezultă că o bună parte din variabilitatea scorurilor se pierde inutil. Îmbunătăţirea sistemului de scorare, prin acordarea unui punct pentru fiecare cub cu o culoare omogenă (roşu sau alb), pus în locul potrivit, şi încă un punct pentru feţele bicolore (roşu şi alb) care respectă nu numai locul, ci şi poziţia relativă (înclinarea), ar putea duce la dublarea ambitusului scorurilor posibile şi deci la o forţă discriminativă dublă. Adăugarea a încă doi itemi (unul foarte uşor, pentru intrarea în probă şi unul foarte greu, pentru a ridica pragul de ieşire din probă), reconsiderarea timpului de aplicare şi al manierei de acordare a bonificaţiilor de timp ar fi încă două surse de mărire a sensibilităţii testului, cu menţiunea că toate aceste modificări reiau de la zero problemele construcţie, de etalonare, de determinare a validităţii şi fidelităţii probei, pentru că, de fapt, aceasta înseamnă crearea unui test nou, ce trebuie studiat ca atare.
- Tipul de norme utilizate în etalonare este a treia caracteristică ce poate contribui la ameliorarea semnificativă a sensibilităţii unui test. Selecţia acestora depinde de precizia dorită în utilizarea testului. Dacă la normele exprimate în centile se poate determina în principiu fiecare punct percentil (caz foarte rar, pentru că testul ar trebui să fie foarte lung şi eşantionul foarte extins), sau măcar punctele decile, la care se mai adaugă percentilele 3, 5, 25, 75, 95 şi 97, dezavantajul constă în inegalitatea unităţilor de măsură (mai mici pe

porţiunea centrală a benzii şi mai lungi spre extreme). În consecinţă, rezultatele unei baterii de teste etalonate în centile nu pot fi adunate şi integrate, prin medie, într-un scor unic. Cuartilele dau o împărţire şi mai grosieră (4 clase), foarte largă, nepermiţând discriminările de fineţe. Testele moderne sunt etalonate în note z (standardizate), sau în norme derivate din acestea, cum sunt notele C, T, Hull, stanine, stens etc. Dintre acestea, foarte utile sunt notele T (cu media de 50 şi abaterea standard de 10), dar şi scalele de QI (cu media 100 şi abaterea standard de 15 sau 16). Ca o regulă vom menţiona că, cu cât o scală va avea mai puţine intervale (11, 10, 7, 5, 3), cu atât mai mult se va produce o pierdere de varianţă şi deci o scădere de sensibilitate a scorurilor testului. Dacă un test de memoria cuvintelor este etalonat în stanine şi clasa 5, la 10 ani are, să zicem, scorurile 48-52, înseamnă că diferenţa de performanţă de 5 puncte dintre subiecţi este practic anulată: fie că scorul este 48 sau este 52, se obţine aceeaşi stanină 5, aşadar sensibilitatea discriminării va avea de suferit. Aceasta nu înseamnă însă că scalele cu mai puţine intervale nu sunt foarte utile (şi utilizate) în practică: multe baterii au subteste foarte diferite ca putere de discriminare şi atunci se alege o scală cu 9-10-11 trepte standardizate (stanine, stens sau note C), care integrează bine toate subtestele în baterie. Unii psihologi s-au acomodat mai bine cu proprietăţile matematice ale unui tip de scală şi îl preferă în virtutea obişnuinţei. Esenţialmente acordajul dintre scorurile la test şi gradul de extensie al scalei utilizate trebuie să ne conducă la alegerea unor scale foarte discriminative când testul are mai mulţi itemi (peste 30-40) eterogeni, nu din punctul de vedere al conţinutului sau al performanţei măsurate, ci al gradului lor de dificultate. Testele ce acoperă multe vârste, diferenţiate între ele prin performanţe ce au ambitusul amplitudinii împrăştierii mare (între 40 şi 160 la testele de inteligenţă, sau chiar mai mari) au nevoie de scale discriminative, celelalte putând opera, pentru scopuri practice şi de cercetare, cu scale mai puţin extinse, chiar dacă aceasta înseamnă scăderea puterii discriminative a testului. Pentru subtestele bateriei sale, Wechsler a optat pentru scale standardizate de 19 trepte.

Cum dificultatea şi sensibilitatea unui test depind de mărimea şi de caracteristicile lotului particular de subiecţi la care aplicăm testul, înseamnă că obligatoriu trebuie să determinăm *valorile tendinţei centrale* pentru a determina gradul de suprapunere ale acestora cu lotul de eşantionare, indicat de manualul testului. Când media grupului nostru este semnificativ diferită de cea a populaţiei de etalonare din manual înseamnă că testul este prea uşor sau prea greu în raport cu aceasta şi concluziile trase, prin raportarea valorilor brute la etalonul standard, pot fi greşite. Dacă abaterea standard indică o valoare mult diferită (mai

mică sau mai mare decât cea a lotului normativ), înseamnă că lotul nostru este fie prea omogen, fie prea eterogen în raport cu eşantionul standard. În prima situaţie forţa de discriminare a testului descreşte, în a doua ea creşte, dar în ambele va trebui avută în vedere construcţia unui etalon mai adecvat pentru populaţia respectivă.

„Trebuie observat că nici un test nu poate fi atât de sensibil încât totdeauna, două persoane deosebite în privinţa variabilei măsurate de test să obţină scoruri diferite", afirmă Monica Albu (1998, p. 189). Aceasta pentru că, în timp ce trăsăturile psihice sunt variabile continue, scorurile la test sunt variabile discontinue (discrete) şi de aceea o corespondenţă biunivocă între variabilă şi scorul la test este un fapt imposibil de realizat practic. Scorul fiind afectat de diverse tipuri de erori (generate de testul însuşi, de subiect sau de situaţia de examinare), este posibil ca aceeaşi variabilă, măsurată la acelaşi individ, să aibă mai multe valori de scor. Întrucât performanţa la un test se raportează la un număr de intervale mai mic decât numărul valorilor de scor pe care el le poate înregistra, înseamnă că indivizii plasaţi pe acelaşi interval de scor nu sunt discriminaţi. Deci un test va fi cu atât mai discriminativ cu cât va separa mai bine indivizii „care au valori ale variabilei măsurate de test plasate în acea zonă a axei sale în care există un număr suficient de intervale de lungimi mici" (Albu, op. cit., p. 190).

3.3. Construcţia unui test psihologic

Analiza de itemi nu se face numai în sensul determinării puterii discriminative a unui test, ci şi în contextul mai general al construirii unui test bun. Deoarece construcţia testelor a devenit în acelaşi timp o ştiinţă şi o artă, ce implică o tradiţie şi experţi în domeniu, nu vom aborda această problemă complexă, prezentând doar principiile generale care stau la baza acestei activităţi.

Elaborarea unui test începe (vezi figura 3.1) cu decizia referitoare la ce fel de informaţie vrem să obţinem prin utilizarea sa. Itemii de genul adevărat-fals vor fi utilizaţi doar când vrem să evaluăm judecăţi absolute, în rest fiind preferabili cei cu alegeri multiple. Şi în acest caz constructorul trebuie să ia în calcul probabilitatea ca cineva să dea răspunsul din întâmplare (prin şansă). Pentru scalele de atitudini sau de opinii, cel mai popular este formatul dat de scala Likert (între acord total şi dezacord total de interpun 3-7 trepte, dintre care treapta intermediară reprezintă neutralitatea). Un alt format este scala în 10 puncte (metoda scalării categoriilor) folosită în *ratingul* unor comportamente, în care punctele de început şi de sfârşit sunt foarte clar definite. *Tehnica Q-Sort,* folosită în studiul personalităţii, cere unui subiect să facă evaluări asupra măsurii în care anumiţi itemi îl descriu pe el sau pe

altcineva (a se vedea capitolul ultim, despre crearea şi statisticile scalelor psihologice). După ce au fost creaţi şi li s-a stabilit formatul, itemii vor fi administraţi unui grup iar caracteristicile lor vor fi apreciate printr-o evaluare sistematică a dificultăţii şi discriminabilităţii itemilor, prin determinarea curbelor caracteristice ale acestora, prin studierea consistenţei interne sau a validităţii relative la criteriu etc. Abia după aceasta se face definitivarea formei finale a testului care va fi aplicat, urmând procedurile de standardizare stabilite (condiţii de aplicare, instructaj, limite de timp şi mod de prelucrare a rezultatelor). Analiza tehnică a testului (fidelitate, validitate şi etalonare) încheie un ciclu care se poate relua de câte ori constructorul doreşte să-şi perfecţioneze sau să reetaloneze testul, să-l îmbunătăţească din punctul de vedere al fidelităţii sau al validităţii prin eliminarea sau adăugarea de itemi.

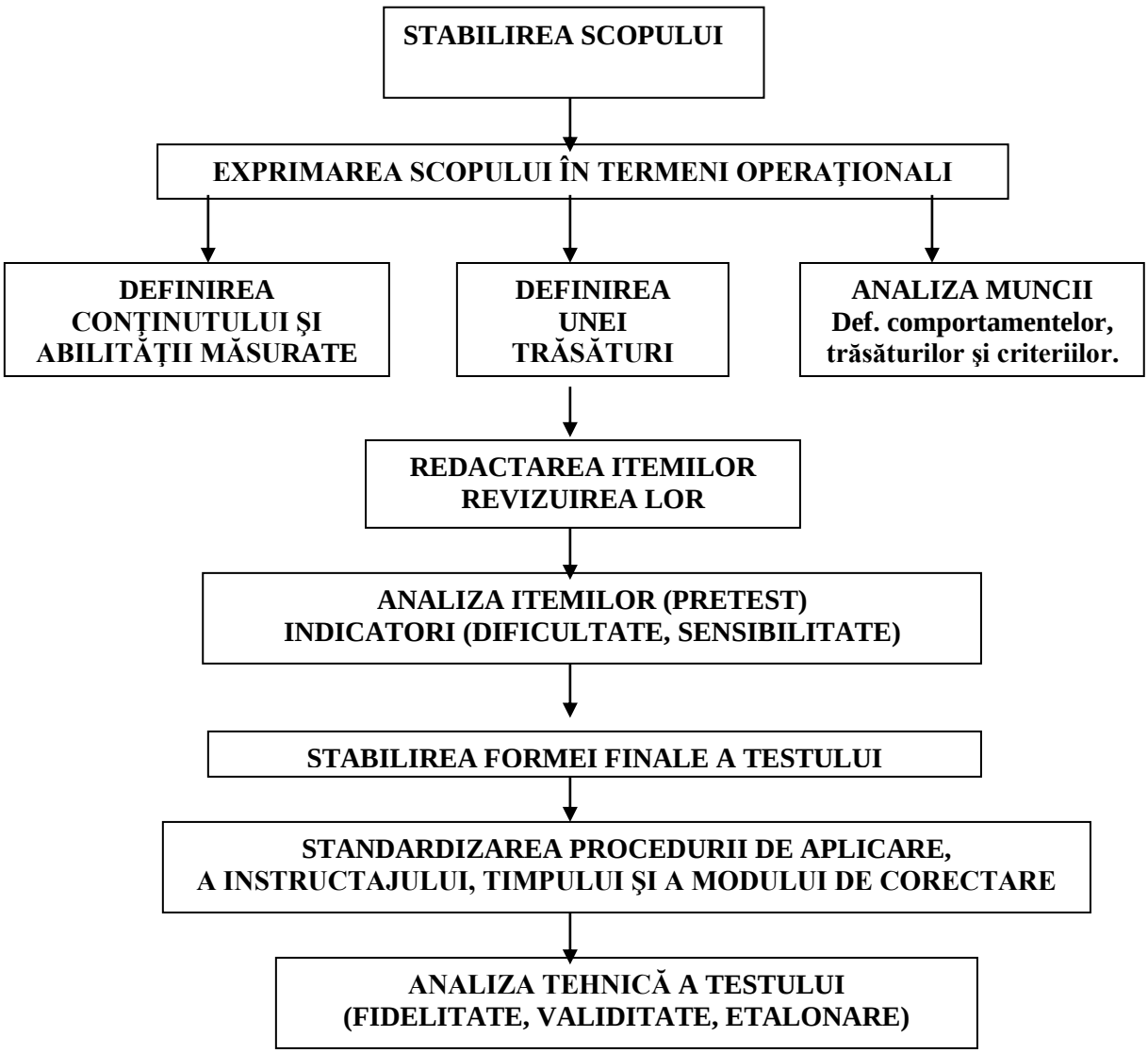
STABILIREA SCOPULUI
EXPRIMAREA SCOPULUI ÎN TERMENI OPERAŢIONALI
DEFINIREA CONŢINUTULUI ŞI ABILITĂŢII MĂSURATE
DEFINIREA UNEI TRĂSĂTURI
ANALIZA MUNCII Def. comportamentelor, trăsăturilor şi criteriilor.
REDACTAREA ITEMILOR REVIZUIREA LOR
ANALIZA ITEMILOR (PRETEST) INDICATORI (DIFICULTATE, SENSIBILITATE)
STABILIREA FORMEI FINALE A TESTULUI
STANDARDIZAREA PROCEDURII DE APLICARE, A INSTRUCTAJULUI, TIMPULUI ŞI A MODULUI DE CORECTARE
ANALIZA TEHNICĂ A TESTULUI (FIDELITATE, VALIDITATE, ETALONARE)

Figura 3.1. *Algoritmul de construire a unui test (după Havârneanu, 2000, p. 130).*

CAPITOLUL 4

FIDELITATEA TESTELOR

4.1. Problematica generală a fidelităţii testelor

Testele sunt concepute ca instrumente de măsură perfecţionate, în consecinţă ele trebuie să aibă calităţile psihometrice presupuse de acest fapt. Dacă pentru lungime, masă, volum, timp sau temperatură există atât instrumente perfecţionate de măsură, cât şi unităţi metrice bine definite, inteligenţa, memoria, personalitatea, motivaţia, iubirea sau sănătatea, agresivitatea, toleranţa la frustrare sunt caracteristici psihologice foarte complexe, ce nu pot fi văzute şi atinse în sens fizic, deci greu cuantificabile.

Cu toate problemele pe care măsurarea le pune în psihologie, preocuparea pentru acurateţea instrumentelor utilizate în sens metric a generat standarde extrem de exigente şi de sofisticate. Plecând de la Spearman (1904), Thorndike (1904), Kuder şi Richardson (1937) şi până la Cronbach (1972, 1989) şi Bentler (1990, 1991), a existat o linie de evoluţie continuă în dezvoltarea unor teorii tot mai elaborate despre fidelitatea probelor psihometrice. Teoria clasică asupra fidelităţii testelor este construită în jurul *erorii de măsurare* care plecă de la ideea că fiecare persoană testată ar avea un scor „adevărat", care ar fi obţinut dacă această eroare nu ar exista. Deci scorul observat (O) este alcătuit din scorul adevărat (A), la care se adaugă eroarea de măsurare (E):

$$O = A + E$$

O altă supoziţie a teoriei clasice a testului este că această eroare este aleatoare, deci scorul adevărat va fi media unei distribuţii în care dispersia reprezintă erorile aleatoare de măsurare. Cum aceste distribuţii pot avea dispersii largi sau înguste, înseamnă că şi erorile de măsurare vor fi mai mari sau mai mici. În felul acesta teoria clasică a folosit abaterea standard a erorilor ca măsură de bază a *erorii de măsurare*, numită şi SEM[1].

Fidelitatea unui set de scoruri este exprimată de un număr zecimal cuprins între 0,00 şi 1,00, indicând absenţa, respectiv fidelitatea perfectă. Deoarece nu poate fi determinată direct, fidelitatea se estimează prin analiza efectelor variatelor condiţii de administrare şi a conţinutului testului asupra scorurilor finale. Ea va fi influenţată doar de schimbările nesistematice care vor avea diferite efecte asupra celor examinaţi. Fiecare din multiplele metode de estimare a fidelităţii va lua în calcul diferitele condiţii ce pot produce asemenea

[1] *Standard Error of Measurement*, adică eroarea standard a măsurătorii.

schimbări nesistematice în scorul testului, ce vor afecta în consecinţă mărimea erorii de măsurare. În funcţie de condiţiile concrete în care a fost administrat testul şi în funcţie de ceea ce acesta măsoară, se calculează unul sau mai mulţi coeficienţi de corelaţie ca o aproximare a fidelităţii testului.

4.2. Fidelitatea test-retest

Numit şi *coeficient de stabilitate*, acesta este obţinut prin corelarea scorurilor obţinute de un grup de persoane la o administrare a testului cu scorurile înregistrate la o administrare ulterioară a acestuia, după un interval de timp. Procedeul încearcă determinarea erorilor legate de condiţiile de aplicare. Cum testul aplicat este acelaşi, inconstanţa scorurilor nu poate fi imputată itemilor săi, ci diferenţei dintre condiţiile de aplicare, cu atât mai mari cu cât intervalul dintre test–retest este mai lung (luni sau ani). Aceasta presupune însă că trăsătura măsurată este ea însăşi stabilă în timp, deci procedeul nu este adecvat pentru unele probe (pentru testele proiective de exemplu, dar nici pentru chestionarele de motivaţii, opinii, dispoziţii afective sau sănătate).

Deoarece pot interveni efectele practicii (unele abilităţi se îmbunătăţesc prin exerciţiu) sau ale învăţării (conţinuturile testului pot fi memorate spre a fi rezolvate ulterior), se pune problema alegerii atente a intervalului dintre cele două examinări. Aceste efecte sunt mai accentuate pentru intervalele scurte (ore sau zile), dar un interval mai lung are dezavantajul de a produce modificări chiar în structura aptitudinii (efectul de creştere sau de maturare, foarte evident mai ales la testele educaţionale). Intervalul optim pentru retest pare a fi de câteva săptămâni până la o lună sau, unde este posibil, se pot calcula coeficienţi de fidelitate pentru intervale de timp diferite (săptămâni, luni sau ani).

4.3. Coeficientul formelor parale

Memorarea testelor nu dă o eroare sistematică, deoarece proporţia itemilor pe care subiecţii şi-i reamintesc după o perioadă este diferită de la individ la individ, ceea ce produce descreşterea corelaţiei test-retest. Pentru depăşirea acestei surse de eroare se poate utiliza procedeul formelor paralele prin calcularea aşa-numitului *coeficient de echivalenţă*, un alt indicator al fidelităţii. Pentru a măsura acelaşi atribut o formă paralelă a unui test trebuie să fie construită în acelaşi mod ca şi prima variantă. Ele vor fi considerate forme paralele doar pentru că utilizează itemi formulaţi diferit, dar procedeul de generare şi selecţie a acestora pentru un anumit nivel de dificultate este acelaşi, deoarece ambele trebuie să măsoare acelaşi construct, în aceeaşi manieră. Formele paralele pot fi aplicate chiar şi în aceeaşi zi, caz în care

singura sursă de diferenţă dintre scorurile la cele două forme este eroarea aleatoare a diferenţei dintre itemii testului.

O procedură mai rafinată utilizată este aceea de a aplica unei jumătăţi din populaţie forma A a testului şi celeilalte forma paralelă B, urmând ca după o perioadă de timp aceluiaşi eşantion să i se aplice tot ambele forme, dar de data aceasta primei jumătăţi forma B iar celei de a doua jumătăşi forma A. Coeficientul de corelaţie ce rezultă se numeşte *coeficient de stabilitate şi echivalenţă*, pentru că ia simultan în calcul, ca surse de eroare, itemii testului şi factorul timp.

4.4. Coeficienţii de consistenţă internă

Din cauza costurilor de construcţie ridicate, nu toate testele au forme echivalente, de aceea se foloseşte o cale mai puţin directă de determinare a fidelităţii: cea a consistenţei interne, ce include metoda înjumătăţirii (*split–half*) a lui Spearman, formulele Kuder-Richardson şi coeficientul alpha (α) al lui Cronbach, care nu trebuie consideraţi echivalenţi cu coeficienţii de fidelitate obţinuţi prin test–retest sau prin forme paralele.

a. Metoda split–half

În engleză, *to split* înseamnă „a despica”, iar *half*, „jumătate”, de unde şi numele metodei, care poate fi tradus prin „înjumătăţire”. Într-un test omogen, toţi itemii măsoară (în diverse grade) acelaşi construct, deci este posibil să construim două jumătăţi relativ echivalente pentru a le pune ulterior în corelaţie. Cele două jumătăţi pot fi obţinute în diverse maniere, cum ar fi: prima parte a testului şi a doua sa parte, sau subteste ce ar rezulta din reunirea tuturor itemilor cu număr par şi cu număr impar (tehnica *par–impar*), sau oricare alt procedeu, chiar şi aleator, de a genera jumătăţi. Problema tehnică spinoasă este că acestea trebuie să fie echivalente, ceea ce în primul exemplu nu se întâmplă: în cele mai multe teste de aptitudini, itemii dificili sunt plasaţi în a doua parte a probei. Chiar şi tehnica par–impar este aplicabilă doar parţial pentru că, deşi în multe teste itemii sunt aranjaţi în ordinea crescătoare a dificultăţii, sunt greu de găsit câte doi itemi perfect echivalenţi din acest punct de vedere, pentru a putea compune cele două jumătăţi. În plus, în unele teste, anumiţi itemi sunt folosiţi pentru a-i introduce pe alţii, adică nu sunt independenţi (itemi legaţi).

Deci metoda split–half presupune analiza de itemi şi calculul mediei şi a abaterii standard, valori care trebuie să fie aproximativ egale pentru fiecare jumătate, pentru a verifica faptul că ele sunt echivalente. Cum corelaţia dintre două seturi de scoruri este mai mică pentru seturile mai scurte, fidelitatea ce rezultă pentru testul în ansamblul său, plecând de la

jumătăţile sale, poate fi estimată prin corecţia pe care „*formula de profeţie*" a lui Spearman-Brown o propune:

$$R = \frac{2r}{1+r}$$

De exemplu, dacă corelaţia dintre jumătăţi este de r = 0,64, fidelitatea testului în ansamblul său va fi R = 2·0,64/(1+0,64) = 0,78. Această formulă are un efect substanţial pentru corelaţiile medii, dar unul mai mic pentru cele extreme.

b. Metoda Kuder-Richardson

În 1937, Kuder şi Richardson au dezvoltat metode de evaluare a fidelităţii dintr-o singură aplicare a unui test, ceea ce a însemnat un mare progres, deoarece aceasta nu mai depindea de maniera arbitrară în care se făcea înjumătăţirea testului. Scorarea separată a jumătăţilor crea o altă problemă, pe lângă cea legată de dificultatea obţinerii unor forme realmente echivalente. Formula de mai jos se poate aplica doar testelor ai căror itemi sunt scoraţi dihotomic (zero sau unu, adică fals sau adevărat):

$$KR_{20} = R = \frac{N}{N-1}\left(\frac{s^2 - \sum pq}{s^2}\right)$$

unde *KR* sunt iniţialele celor doi psihologi, *R* este fidelitatea estimată, *N* numărul de itemi ai testului, s^2 varianţa scorului la test în ansamblul său, *p* este proporţia (calculată pentru fiecare item în parte) în care un item este rezolvat corect, *q* este complementul acestei proporţii, adică *(1 – p)*, Σpq este suma produselor pq pentru fiecare item al testului. Studiind formula, vom vedea că partea ei din dreapta comportă o analiză din care rezultă că pentru a avea o fidelitate mai mare ca zero, varianţa testului trebuie să fie mai mare ca suma varianţelor individuale ale fiecărui item ($s^2 > \Sigma pq$). Acest lucru este posibil doar în cazul în care itemii, în calitatea lor de măsură a aceluiaşi construct, sunt intercorelaţi.

Pentru situaţia când itemii celor două jumătăţi sunt aproximativ egali ca nivel de dificultate (cazul echivalenţei), cei doi autori au propus o altă formulă, mai uşor de calculat, dar care nu se poate aplica jumătăţilor neechivalente, întrucât le va subevalua fidelitatea:

$$KR_{21} = \frac{N}{N-1}\left[1 - \frac{\bar{x}\left(1 - \frac{\bar{x}}{N}\right)}{s^2}\right]$$

unde toate notaţiile sunt aceleaşi ca în formula precedentă, iar $\bar{x}$ reprezintă media scorurilor totale la test.

c. Coeficientul alpha (α) al lui Cronbach

Formula KR_{20} este o foarte valoroasă procedură de estimare a consistenţei interne a unui test. Cu toate acestea, există situaţii în care ea nu este potrivită, iar acestea apar atunci când testul nu a fost scorat în termeni de adevărat–fas, sau zero–unu, deoarece formula se bazează pe cunoaşterea proporţiei în care subiecţii au rezolvat corect fiecare item. Unele chestionare de opinii, atitudini sau valori sunt elaborate nu în termeni dihotomici, ci presupun o scară a gradului de acord–dezacord sau atracţie–respingere. În această situaţie, Cronbach (1951) a elaborat o procedură mai generală de estimare a fidelităţii, aşa-numitul *coeficient alpha*, a cărui formulă este:

$$R = \alpha = \frac{N}{N-1}\left(\frac{s^2 - \sum s_i^2}{s^2}\right)$$

Se observă o mare similitudine cu formula KR_{20} chiar şi în notaţii, termenul nou fiind s_i^2. Termenul Σs_i^2 înlocuieşte în această formulă Σpq şi el reprezintă suma varianţelor individuale ale fiecărui item. Singura diferenţă constă deci în maniera în care este exprimată varianţa. Deoarece s_i^2 exprimă varianţa itemilor care nu sunt rezolvabili doar prin „da/nu”, coeficientul alpha este considerat a fi soluţia cea mai generală pentru determinarea fidelităţii pentru consistenţa internă.

Kaplan şi Saccuzzo (1993) sesizează foarte întemeiat că toate măsurile care evaluează consistenţa internă, determină de fapt gradul în care fiecare item diferit măsoară aceeaşi trăsătură sau abilitate. Acest lucru presupune implicit că testul este omogen, în caz contrar acesta neavând consistenţă internă. Pentru testele neomogene, procedeul cel mai indicat este analiza factorială, prin care se vor putea subîmpărţi itemii pe grupe omogene, subtestele ce rezultă având fiecare o consistenţă internă ridicată, dar fiind relativ independente unul în raport cu celelalte, ca în cazul subscalelor testului de prsonalitate 16PF Cattell.

Aiken (1997) sesizează de asemenea că cele trei procedee de determinare a fidelităţii, amintite anterior, supraestimează valoarea acesteia pentru testele în care este implicată viteza. În acest caz procedurile de apreciere a fidelităţii trebuie modificate, recomandarea făcută fiind aceea de a administra cele două jumătăţi ale testului în momente diferite, dar cu limită de timp egală. După aceasta se calculează fidelitatea, operându-se corecţia ei prin formula Spearman–Brown.

d. Fidelitatea interscoreri

Majoritatea testelor de aptitudini, de performanţă şi de personalitate au proceduri de scorare standardizate, aşa că nu va fi nici o problemă în privinţa scorurilor obţinute de către persoane diferite, devreme ce aceste proceduri de apreciere pot fi încredinţate maşinilor automate de scorare sau computerului.

Tehnicile proiective, judecăţile evaluativ–apreciative, evaluarea unor produse complexe ale activităţii (desene, mostre de scris, obiecte manufacturate etc.) sau ratingul personalităţii include, prin natura lucrurilor, o doză crescută de subiectivitate. În acest caz trebuie să determinăm gradul de intervenţie în apreciere al subiectivităţii, determinând *fidelitatea interscoreri* sau interevaluatori prin calculul corelaţiei dintre două seturi de scoruri acordate de evaluatori diferiţi, unui număr determinat de examinaţi. Se poate apela şi la metoda „mai mulţi examinatori – un singur examinat", sau „mai mulţi examinatori – mai mulţi examinaţi", procedee care culeg *coeficienţi de fidelitate intraclasă*, sau *de concordanţă* (cum ar fi coeficientul tau al lui Kendall, vezi Radu *et al*., 1991), pentru care există programe speciale de calculator, cum ar fi SPSS.

4.5. Factorii care intervin în determinarea fidelităţii

Pentru ca estimarea fidelităţii să fie cât mai precisă, trebuie îndeplinite câteva condiţii (Traub, 1944, apud Albu, 1998):

- Eşantionul să fie cât mai mare, pentru a reduce eroarea standard a repartiţiei, care este invers proporţională cu rădăcina pătrată din numărul subiecţilor ce compun eşantionul.
- Eşantionul să fie reprezentativ pentru populaţia căreia îi este destinat testul, având acelaşi grad de eterogenitate, deoarece omogenitatea diminuează valoarea coeficientului de fidelitate.
- Măsurătorile să fie independente între ele, astfel încât măsurătoarea de la un examinator să nu o influenţeze pe a altuia, iar dacă examenul este colectiv, să nu se poată trişa prin copiere. Independenţa cere ca persoanelor supuse la test–retest să nu li se dea informaţii despre rezultatele examinării precedente, iar itemii să nu se condiţioneze reciproc.
- Toate aspectele de procedură în test şi retest sau la formele paralele trebuie să fie identice.

Creşterea nivelului de fidelitate a unui test până la limita dorită sau cerută de situaţia de utilizare concretă este posibilă prin creşterea numărului de itemi (care trebuie să fie de acelaşi format şi să măsoare aceeaşi trăsătură sau acelaşi construct)[2]. Decizia aceasta angajează un proces lung şi costisitor, pentru că testul nou generat trebuie reevaluat de la început şi uneori se dovedeşte a fi sub nivelul de fidelitate aşteptat. De asemenea, crescându-i lungimea, testul devine mai greu de aplicat şi de scorat. Manipulând formula de profeţie a lui Spearman–Brown se calculează un indice de multiplicare a numărului de itemi pentru a atinge fidelitatea dorită: un test de 20 de itemi trebuie să ajungă la 56 de itemi pentru a-i creşte fidelitatea de la 0,87 la 0,95 (Kaplan şi Saccuzzo, op. cit., p. 127).

Ca o concluzie a acestor consideraţii despre fidelitatea testelor, se relevă faptul că aceasta este mai bună pentru teste unidimensionale şi cu număr mai mare de itemi, cât şi pentru testele cognitive şi că ea angajează un studiu analitic al itemilor (*analiza de itemi*) pentru a le determina forţa de discriminare. În multe situaţii analiza factorială este metoda cea mai eficientă prin care se pot construi subteste omogene şi unidimensionale. „Fidelitatea este una din fundamentările de bază ale cercetărilor asupra comportamentului. Dacă un test nu este fidel, nu va fi posibil să demonstrăm că el are vreun înţeles", afirmă Kaplan şi Saccuzzo (op. cit., p. 131). Deci, deşi validitatea pare a fi o caracteristică psihometrică mai importantă, din punct de vedere tehnic studiul ei nu poate începe cu teste care nu îşi dovedesc o fidelitate minimă, acceptabilă, care depinde de scopul în care acestea vor fi utilizate.

[2] Vezi capitolul următor.

CAPITOLUL 5

FACTORI CARE AFECTEAZĂ FIDELITATEA TESTELOR

5.1. Întinderea diferenţelor individuale

Deoarece estimarea fidelităţii rezultă din calculul corelaţiei r Pearson limitele statistice ale acestuia se repercutează şi asupra coeficientului de fidelitate. Astfel, r este calculat pentru varianţa totală a testului – ceea ce este o condiţie necesară, dar nu şi suficientă pentru a determina fidelitatea. Atunci când se produce o reducere sau o creştere a ambitusului scorurilor individuale, corelaţia dintre variabile (şi implicit fidelitatea testului) scade sau creşte în mod artificial, diminuând şansa de a determina corect fidelitatea sa. Astfel, dacă în faza de studiu pilot se utilizează un eşantion care are o varianţă mai mică decât populaţia generală, fidelitatea testului va fi subestimată (va fi mai mică decât dacă eşantionul ar fi fost reprezentativ). Aplicarea testului pe o categorie populaţională foarte omogenă (clase de elită sau de subdotaţi intelectual, clase vocaţionale etc.) ar putea furniza o corelaţie mai mică decât cea existentă la clasele normale, unde extremele (subdotaţi, supradotaţi intelectual) sunt reprezentate normal.

Varianţa – şi implicit fidelitatea – poate fi crescută artificial prin folosirea de eşantioane cu grad mare de eterogenitate. Agregarea într-un eşantion comun a unor clase de elită (supradotaţi), normale şi de integrare (subdotaţi) accentuează mărimea corelaţiei, ca şi reunirea în acelaşi eşantion a unor niveluri de pregătire şcolară aflate la mare distanţă unele de altele. Pentru situaţia în care eşantionul a fost prea omogen, diminuând astfel varianţa totală a scorurilor la test, şi deci fidelitatea sa, Magnuson a propus o formulă de corecţie:

$$r_{uu'} = \frac{1 - \sigma_x^{\,2}\left(1 - r_{xx'}\right)}{\sigma_u^{\,2}}$$

în care $r_{uu'}$ este fidelitatea estimată pentru noul eşantion, σ_x^2 este varianţa noului eşantion iar σ_u^2 este fidelitatea calculată între vechiul şi noul eşantion.

Pentru utilizatorul unui test precauţiile legate de întinderea diferenţelor individuale sunt două:

- utilizarea tabelelor de norme, care aduc distribuţiile la „un numitor comun” prin raportarea la notele standardizate z;

- studiul pilot al fidelităţii pe propriul eşantion de lucru pentru a ne asigura că populaţia noastră corespunde ca raport de omogenitate/eterogenitate populaţiei pe care s-a determinat fidelitatea raportată în manualul testului.

5.2. Lungimea unui test

Teoria eşantionajului demonstrează faptul că, cu cât un eşantion este mai mare, cu atât mai mult estimarea caracteristicilor populaţiei din care acesta a fost extras este mai exactă. În mod similar, cu cât numărul de itemi ai unui test este mai mare, cu atât mai bine este măsurat constructul sau domeniul investigat, şi aceasta deoarece suma erorilor aleatorii tinde tot mai mult spre zero.

Relaţia dintre fidelitatea şi lungimea testului este exprimată de *formula de profeţie* a lui Spearman-Brown, care ne va arăta cu cât creşte precizia estimarii prin modificarea numărului itemilor într-o anumită proporţie K:

$$r_{xx'} = \frac{K \cdot r_{jj'}}{1+(K-1)r_{jj'}}$$

în care $r_{xx'}$ este fidelitatea expectată prin lungirea testului, iar $r_{jj'}$ este fidelitatea calculată pentru testul iniţial. Această formulă poate fi utilizată în două feluri:

- dacă iniţial testul avea 20 de itemi şi dorim să-l aducem la 45 de itemi, K reprezintă raportul 45/20 = 2,25. Plecând de la fidelitate iniţială de 0,83, fidelitatea expectată prin lungirea testului va ajunge la: (2,25·0,83)/[(1+2,25-1)·0,83] = 1,8675/(1+1,0375) = 1,8675/2,0375 = 0,92;
- invers, dacă vrem să ştim câţi itemi trebuie să adăugăm pentru a atinge valoarea dorită a fidelităţii, de exemplu creşterea de la 0,83 la 0,95 (pentru a putea ajunge în zona în care putem lua decizii care privesc destinul unei persoane), maniera de lucru presupune mai întâi izolarea lui K, plecând de la formula anterioară.

$$K = \frac{r_{xx'}(1-r_{jj'})}{r_{jj'}(1-r_{xx'})}$$

În cazul nostru: $K = \frac{0,95(1-0,83)}{0,83(1-0,95)} = \frac{0,95 \cdot 0,17}{0,83 \cdot 0,05} = \frac{0,16 \cdot 5}{0,04 \cdot 5} = 3,89$

- aceasta înseamnă că, pentru a avea creşterea de fidelitate expectată, testul trebuie mărit de 3,89 ori şi deci de la 20 de itemi el va ajunge la 78 (20·3,89 = 77,8).

Formulele de mai sus pot fi folosite şi în sensul scurtării unui test prea lung (cu scăderea de rigoare a fidelităţii sale) dar fie într-o situaţie, fie în alta, itemii trebuie să aibă acelaşi conţinut şi acelaşi grad de dificultate, fiind consistenţi cu itemii de plecare. Ori aceasta înseamnă parcurgerea prealabilă a fazei analizei de itemi, căci itemii foarte diferiţi ca nivel de dificultate sau ca şi conţinut nu vor avea o bună corelaţie cu cea deja existenţi, scăzând omogenitatea de ansamblu a testului.

Metoda Spearman-Brown nu poate preciza care sunt caracteristicile itemilor ce vor fi adăugaţi – în termeni de format şi de conţinut al acestora – pentru a face să crească fidelitatea testului până la o precizie antecalculată acceptabilă. Dacă în testele de aptitudini itemii ce vor fi adăugaţi vor fi definiţi în acelaşi fel ca şi ceilalţi, pentru a fi corelaţi cu constructul măsurat, în testele educaţionale ei vor viza aceleaşi obiective pedagogice ca şi testul iniţial. Acesta nu este un lucru dificil, căci itemii paraleli se creează uşor dacă am folosit din start tehnica specificării domeniului. Principalul inconvenient al metodei este utilizarea unui demers empiric în crearea setului de itemi. Fidelitatea poate fi mai bine ameliorată atunci când un test a fost construit după demersul criterial, în care caracteristicile itemilor trebuie să fie bine cunoscute.

5.3. Dificultatea testului

Corelaţia dintre două teste tinde să fie maximă doar atunci când distribuţiile celor două variabile au acelaşi tip de asimetrie. Aşa cum am arătat în paginile anterioare, în fazele testării formării unei deprinderi aceasta poate avea o asimetrie pozitivă în faza iniţială (testul este prea greu) şi una negativă în faza finală (testul devine prea uşor), trecând printr-o fază intermediară unde există simetrie a distribuţiei scorurilor. Asimetriile fiind în direcţii opuse, cea mai mică fidelitate a testului va fi corelarea dintre faza iniţială şi faza finală, după care corelarea dintre faza intermediară şi celelalte două faze, iniţială sau finală.

Schimbarea formei distribuţiei poate fi principala cauză a unei fidelitaţi mici, prin faptul că un test a devenit mai uşor la o a doua aplicare, dând o curbă asimetrică negativ, ceea ce presupune contaminarea rezultatelor ca efect al învăţării. În consecinţă, verificarea formei distribuţiei scorurilor pentru acele două aplicări devine obligatorie.

5.4. Testele cu limită de timp

Multe teste de cunostinţe sau de aptitudini conservă primatul gradării dificultăţii itemilor, de la cei mai uşori la începutul testului (pentru creşterea încrederii în sine), la cei mai dificili spre sfârşitul acestuia (pentru a creşte puterea diagnostică a instrumentului în

zona vârstelor mari ori a supradotării). Dacă ele au o limită de timp, itemii dificili nici măcar nu vor fi abordaţi de majoritatea subiecţilor, fiind scoraţi cu zero şi la prima aplicare, şi la cea de a doua. Acest fapt va crea o creştere artificială a corelaţiei prin care se pune în evidenţă fidelitatea, deoarece vor intra în corelaţie multe perechi identice de rezultate (0 - 0). „Inflaţia" fidelităţii poate crea imaginea distorsionată a unui test cu itemi ce vor apărea ca mai omogeni decât sunt de fapt în realitate. De fapt, la testele de viteză nu toate formele de determinare a fidelităţii sunt afectate. În timp ce în cazul jumătăţii vii (*split-half*) aceasta va fi afectată (itemii scoraţi zero se distribuie aproximativ egal în cele două jumătăţi ale testului, crescând artificial fidelitatea), consistenţa internă prin indicele alpha al lui Cronbach sau prin metoda test-retest nu va fi afectată semnificativ în acest caz.

Pentru a determina totuşi şi celelalte forme de fidelitate se dau testul şi retestul cu limita de timp convenită de autor, se marchează ultimul item (de aceea itemii trebuie pasaţi în ordine, şi nu „pe sărite") apoi se continuă până la capăt testul, fără limită de timp. Scorurile din prima categorie vor servi pentru determinarea mediilor, a abaterilor standard şi a formei distribuţiei, necesare scopurilor avute în vedere (cercetare, crearea de etaloane), în timp ce scorurile brute obţinute fără limită de timp vor da o mai bună expresie a fidelităţii prin metodele split-half şi alpha al lui Cronbach.

5.5. Fidelitatea şi eroarea de măsurătoare

Deoarece fidelitatea nu exprimă valoarea preciziei măsurătorii în aceleaşi unităţi ca şi scorul total la test (exprimat în note standard), uneori ea este mai greu de interpretat. Acesta este motivul pentru care precizia măsurătorii poate fi indicată şi sub forma unei erori de interpretare a scorului la test, care va fi cu atât mai mică cu cât eroarea de măsurare va fi şi ea mai mică. Această zonă de încredere poate fi determinată în două maniere:

a) se determină *eroarea de măsurătoare* (interval de încredere înăuntrul căruia se află adevăratul scor al unui subiect, pentru niveluri de încredere specificate);

b) se determină *eroare de estimare* (nivelul de încredere al scorului observat dacă subiectul ar fi retestat).

5.5.1. Eroarea standard a măsurătorii (SEM)

Conform teoriei clasice a testului, scorul adevărat al unui subiect se distribuie normal în jurul unei valorii medii, dând o distribuţie pentru care putem determina abaterea standard. Abaterea standard a diferitelor distribuţii rezultate pentru toţi subiecţii grupului în cauză se numeşte SEM (σ_e) adică *Eroarea Standard a Măsurătorii* şi se determină astfel:

$$\text{SEM} = \sigma_e = \sigma_x \sqrt{1 - r_{xx'}}$$

în care: $r_{xx'}$ este coeficientul de fidelitate iar σ_x este abaterea standard de la care s-a plecat în calculul coeficientului de fidelitate $r_{xx'}$. De exemplu, pentru o fidelitate de 0,93 şi o abatere standard de 12, $\sigma_e = 12\sqrt{1-0{,}93} = 3{,}17$. Plecând de la premisa că eroarea de măsurătoare este normal distribuită, 68% din scoruri se vor situa între un interval de $\pm 1\sigma_e$ în jurul scorului adevărat, 95% la $\pm 1{,}96\sigma_e$ iar 99% la $\pm 2{,}58\sigma_e$. Nu ne rămâne decât să definim aceste intervale. Cum noi nu cunoaştem adevăratul scor al subiectului, este preferabil să construim intervalul de încredere în jurul a ceea ce deja cunoaştem, adică nota observată x (obţinută de subiect) şi eroarea standard a măsurătorii:

$$x - z_c\sigma_e \leq A \leq x + z_c\sigma_e$$

în care:

x = scorul observat;

z_c = valoarea critică a lui z pentru $p \leq 0{,}05$ ($z = 1{,}96$) sau $p \leq 0{,}01$ ($z = 2{,}58$);

σ_e = eroarea standard a măsurătorii;

A = scorul adevărat.

Exemplu: scorul x al unui subiect la un test de inteligenţă este de 121, eroarea standard a măsurătorii este de 3,17 şi vrem să construim intervalul de încredere pentru $p \leq 0{,}05$ şi $p \leq 0{,}01$. Pentru prima situaţie avem $121 \pm 1{,}96 \cdot 3{,}17 = 121 \pm 6{,}21$. Deci pentru un $p \leq 0{,}05$, intervalul este [115; 127], adică sunt 95% şanse ca scorul adevărat al subiectului să cadă între 115 şi 127.

Pentru a doua situaţie avem: $121 \pm 2{,}58 \cdot 3{,}17 = 121 \pm 8{,}18$, deci pentru un prag de încredere $p \leq 0{,}01$, intervalul definit este [113; 129], existând 99% şanse ca scorul subiectului să cadă între aceste limite. Se observă de aici faptul că: a. cu cât fidelitatea este mai mare, cu atât mai mult limitele intervalelor de încredere pentru orice nivel de încredere calculat sunt mai mici (mai strânse) şi b. cu cât vrem să ştim cu o precizie mai mare în ce zonă cad scorurile cuiva, cu atât mai larg va fi intervalul de încredere rezultat, şi reciproc.

Aceste intervale de încredere fie sunt adesea neluate în seamă de pracricieni, fie sunt rău interpretate. De fapt noi nu avem nicicum certitudinea că scorul adevărat al subiectului va cădea în intervalul de încredere determinat, căci există doar o probabilitate de a fi aşa şi, în al doilea rând, (şi aceasta pare a fi defiecienţa majoră a acestei metode) intervalele de încredere astfel determinate se sprijină pe postulatul homoscedasticităţii, adică pe prezumţia că eroarea tip este aceeaşi pe toată scara nivelurilor successive ale performanţei. În al treilea rand, corelaţia dintre scorul observat şi cel estimat nu este niciodată una perfectă (Nunnally &

Bernstein, 1994) şi în consecinţă prezicerea scorului adevărat plecând de la cel observat produce fenomenul de regresie spre medie al scorurilor adevărate. Luând în consideraţie această problemă, Glutting, McDermott şi Stanley (1987) au propus o modalitate de construcţie a intervalului de încredere mult mai riguroasă, după formula:

$$\hat{A} = x + r_{xx'}(x - \bar{x})$$

în care Â este scorul adevărat estimat, x scorul observat, $\bar{x}$ media scorurilor observate, iar $r_{xx'}$ este coeficientul de fidelitate.

Eroarea tip a estimării se va calcula şi ea după o formulă modificată:

$$\sigma_{eu} = (\sigma_x \sqrt{1 - r_{xx'}}) r_{xx'}$$

în care σ_x este eroarea standard a distribuţiei de la care s-a calculat $r_{xx'}$ iar $r_{xx'}$ este chiar coeficientul de fidelitate calculat. Pentru exemplul nostru, în care un subiect obţinea 121 la un test de inteligenţă cu media 100 şi cu abaterea standard de 15, testul având fidelitatea de 0,93, determinarea scorului adevărat devine: 100 + 0,93(121 - 100) = 100 + 0,93·21 = 100 + 19,53 = 119,53, rotunjit 120. Calculul erorii standard a estimaţiei este următoarea: (15√1-0,93)·0,93 = 15·0,26·0,93 = 3,69. Intervalul de încredere pentru p = 0,05 este egal cu 1,96·3,69 = 7,23 în jurul scorului adevărat estimat, adică 120 ± 7. Acesta este intervalul [113-127]. Pentru pragul de p ≤ 0,01, eroarea va fi 7,58 - 3,69 = 9,52, iar intervalul va fi 120 ± 9,52, adică ≈ [110-130]. Aşa cum se observă în raport cu scorul observat de 121 intervalul nu mai este unul simetric în jurul valorii obţinute de subiect: în prima situaţie de la 121 – 113 = 8 puncte de scor, pentru limita inferioară, şi în a doua situaţie sunt 127 – 121 = 6 puncte, pentru limita superioară. Această asimetrie provine din centrarea intervalelor de încredere diferite nu pe scorul observat, ci pe cel estimat ca adevărat, ceea ce face din această procedură cea mai riguroasă metodă de determinare a intervalelor de încredere.

5.5.2. Eroarea tip a estimării

Eroarea tip a estimării se regăseşte de fiecare dată când dorim să calculăm intervalul de încredere a valorii prezise plecând de la o ecuaţie de regresie liniară. Acest tip de eroare se obţine extrăgând rădăcina pătrată din varianţa reziduală, adică acea varianţă a scorurilor care se regăseşte la al doilea test atunci când se ţine cont de primul test. Cum metoda de calcul şi interpretarea erorii de estimaţie sunt mai laborioase, ele nu au fost incluse în capitolul de faţă. Menţionăm de asemenea faptul că în corelaţia liniară – pe care se bazează de regulă determinarea fidelităţii – homoscedasticitatea este presupusă ca o condiţie implicită. Deoarece eroarea de măsurătoare este mai mică la cei care au preponderant reuşite sau

nereuşite (extremele seriei de variaţie a performanţelor), comparaţi cu cei de pe porţiunea de mijloc, Keats şi Lord au propus un model fundamentat pe distribuţia binominală, care permite estimarea erorii tip a măsurătorii indiferent de nivelul scorului subiecţilor. Nici această procedură nu a fost inclusă aici.

5.5.3. Interpretarea fidelităţii

Răspunsul la întrebarea „cât de mare trebuie să fie fidelitatea unui test?" depinde de utilitatea practică care se dă acestuia.

- O fidelitate de 0,70 – 0,80 este suficient de bună atunci când testul este folosit în scopuri de cercetare.
- Se acceptă niveluri scăzute ale fidelităţii atunci când testele se utilizează pentru a lua decizii preliminare şi ridicate pentru decizii finale, sau când ele sunt folosite pentru împărţirea grupului în subgcategorii, pe baza unor diferenţe interidividuale mari.
- Dacă testul serveşte la compararea grupurilor de persoane între ele, coeficienţii de fidelitate de 0,60 – 0,70 sunt suficienţi, dar când testul devine o bază de comparaţie între persoane individuale, fidelitatea lui trebuie să fie de la 0,85 în sus.
- Când se iau decizii importante pe bază de teste, prin care se împart persoanele în categorii, în virtutea unor diferenţe mici (ca în selecţia profesională), fidelitatea acestora trebuie să fie de peste 0,90.
- Când decizia priveşte destinul unei persoane individuale, fidelitatea testului trebuie să fie de cel puţin 0,95.

Testele cognitive, şi în special cele de inteligenţă, au de regulă o fidelitate foarte mare (peste 0,90), în timp ce chestionarele de personalitate rareori depăşesc 0,80. Aplicate colectiv, chiar şi testele cognitive furnizează coeficienţi de fidelitate mai scăzuţi (în jur de 0,80). Testele cotate subiectiv, ce măsoară aptitudini, şi testele de cunoştinţe (educaţionale) rareori depăşesc valori ale fidelităţii de 0,80 (Traub, 1944). Testele cu alegere multiplă, utilizate colectiv, sunt considerate a avea o fidelitate bună când aceasta atinge 0,75.

5.5.4. Generalizabilitatea

Potrivit paradigmei clasice relativă la fidelitate, O = A + E, descrisă în partea introductivă a capitolului de faţă, fidelitatea unui instrument psihometric este în funcţie de întinderea spectrului diferenţelor individuale, de lungimea testului însuşi, de limita de timp acordată şi de dificultatea testului. Cu toate acestea condiţiile de observare şi de măsurătoare

sunt mult mai complexe decât cele enumerate anterior, ceea ce a condus la conturarea unei noţiuni noi şi anume cea de **generalizabilitate**. Aceasta presupune studiul fidelităţii în familii de situaţii similare, ceea ce face ca *scorul adevărat* (A, din formula de mai sus) să fie înlocuit cu termenul de *univers al scorului* aşteptat de la un subiect, într-un anumit ansamblu de condiţii de observare şi de măsurătoare.

Teoria generalizabilităţii a fost elaborată în 1963 de Cronbach, Gleser şi Rajaratnam cu scopul de a reuni într-un concept unic diferitele definiţii ale fidelităţii. Astfel, pentru a cuantifica importanţa fiecărei surse de varianţă dintr-o situaţie de măsurătoare ei au folosit analiza de varianţă. Scorul adevărat îşi lărgeşte înţelesul pentru că ia în calcul toate observaţiile posibile, împreună cu erorile aferente rezultate din fluctuaţiile de eşantionaj legate de momentele de evaluare, de forma itemilor sau de „ecuaţia personală" a fiecărui evaluator în parte. Din această cauză generalizabilitatea este un concept mai cuprinzător decât cel de fidelitate, căci el descrie situaţia de măsurare într-un cadru mai complex şi mai apropiat de realitate. În esenţă aceasta indică măsura în care se poate generaliza un rezultat obţinut în anumite condiţii.

Noţiunea de *univers al scorului* se cheamă astfel pentru că are în vedere fidelitatea scorurilor într-un întreg univers de condiţii care alcătuiesc sau definesc tot atâtea faţete ale planului de observaţie. Iată definiţia dată de Cardinet şi Turneur (1985, p. 23) universului scorului: „Universul scorului unei persoane p, dată ideal, reprezintă media scorurilor persoanei p calculată pe toate observaţiile admisibile. Ori observatorul utilizează scorul observat, sau o funcţie a scorului observat pentru a estima valoarea universului scorului. El generalizează astfel de la eşanion către populaţia de ansamblu."

O paralelă cu fidelitatea se impune şi pentru generalizabilitate. Dacă în primul caz, cu cât corelaţia dintre scorul observat şi cel adevărat este mai mare, şi fidelitatea este mai bună, în generalizabilitate aceasta este cu atât mai mare cu cât scorul observat la un subiect este mai asemănător cu cel pe care el l-ar fi obţinut în ansamblul de condiţii pentru care vrem să generalizăm. Cum universul scorului nu poate fi obţinut direct, ceea ce ne rămâne de făcut este eşantionarea sa. Aici problema cheie este cea legată de rezolvarea unei contradicţii: generalizabilitatea este cu atat mai mare cu cât ţine sub control mai multe faţete ale dispozitivului de măsurare, ceea ce înseamnă sporirea numărului de itemi (de teme) investigate, al numărului de corectori şi de grile de corecţie, dispozitivul devenind astfel foarte costisitor. Ar trebui găsit deci un echilibru între economicitatea şi eficacitatea dispozitivului de măsurare, ori pentru aceasta ar trebui determinată ponderea fiecărei faţete a examinării, doar astfel putând să le reunim într-un model eficace.

Pentru a ţine cont de multitudinea variaţiilor ce se produc între diversele faţete ale unui dispozitiv de măsurare, ca şi de diversele interacţiuni posibile dintre acestea, studiul generalizabilităţii are în mod expres nevoie de analiza de varianţă.

5.5.5. Analiza de varianţă şi planul de optimizare

Studiul generalizabilităţii permite un control crescut al surselor de eroare dintr-un dispozitiv de măsurare (sau dispozitiv observaţional). Dincolo de calculul unui indice de fidelitate al unui univers al scorului, cercetătorul poate detrmina în ce condiţii sau situaţii dispozitivul său va prezenta caracteristicile cele mai bune ale măsurătorii. În forma iniţială a teoriei generalizabilităţii Cronbach, Gleser, Nanda şi Rajaratnam (1971) s-au interesat doar de stabilitatea scorurilor subiecţilor. În psihologie şi pedagogie merită tot atâta interes un numai subiecţii, ci şi itemii testuluo, căci se poate estima stabilitatea diferitelor modalităţi de prezentare sau de evaluare a lor, ca şi diferitele conţinuturi ce fac obiectul testului (măsurătorii) respective. Cardinet şi Tourneur (1985) au definit un procedeu de calcul care permite să se ţină cont în dispozitivul de măsurare respectiv atât de subiecţi, cât şi de itemii utilizaţi. În consecinţă ei au indicat cei patru paşi ai procesului, primii doi fiind o analiză de varianţă, faza a treia studiază diferenţierea iar faza a patra optimizarea.

Mai jos facem o succintă prezentare a acestei proceduri:

1. ***Plan de observaţie***: se procedează la alegerea faţetelor şi a numărului de niveluri ale fiecărei faţete, precizându-se şi interrelaţiile dintre ele.
2. ***Plan de estimare:*** se determină care faţete reprezintă un ansamblu de niveluri finite şi care dintre ele sunt eşantionate aleator sau exhaustiv.
3. ***Plan de măsurare***: se identifică ce faţete sunt legate de planul de măsurare (faţete ale diferenţierii) şi care sunt sursele de eroare ale măsurării (faţetele de instrumentaţie). Cu alte cuvinte abia acum varianţa calculată în faza a doua se atribuie fie varianţei adevărate, fie varianţei erorii, permiţând calculul coeficientului de generalizabilitate şi calculul marjei de eroare aplicabilă scorurilor observate.
4. ***Plan de optimizare***: acesta permite modificarea oricăruia din planurile precedente, ca şi a combinaţiilor dintre ele, cu scopul de a maximiza generalizabilitatea observaţiilor (măsurătorilor). Este faza în care cercetătorul caută acel echilibru dintre precizia măsurătorii şi întinderea universului generalizării: cu cât acesta este mai restrâns, este mai uşor de obţinut măsurători adecvate ale acestuia, fapt care este mai dificil în situaţia inversă.

CAPITOLUL 6

VALIDITATEA

Spre deosebire de fidelitate, care este influenţată numai de erorile de măsurare nesistematice, validitatea unui test este afectată atât de erorile nesistematice, cât şi de cele sistematice (constante). Din această cauză, „un test poate să fie fidel, fără a fi valid, dar nu poate fi valid fără să fie fidel“, afirmă Aiken (1993, p. 94).

Accepţiunea clasică a validităţii se referă la gradul în care un test măsoară ceea ce şi-a propus, ceea ce sugerează că ar exista doar un singur tip de validitate. Actualmente este unanim acceptat faptul că un test poate avea diferite forme de validitate, dependente de scopurile specifice în care el a fost utilizat. De aceea ni se pare de mare actualitate afirmaţia Ursulei Şchiopu (1974, p. 156) potrivit căreia un test este bidimensional, axat pe doi factori, adică pe diagnoză şi pe validarea ei, concepuţi ca axele unui sistem de coordonate: „în acest sistem axa X reprezintă o diagnoză cuantificabilă (test – comportament), axa Y reprezintă criteriul de valabilitate al diagnozei”.

În acest sens poate fi înţeleasă sintagma că *un test se validează continuu*. Făcând o analogie cu justiţia, unde operează prezumţia de nevinovăţie (fără dovezi concrete, cel trimis în instanţă este nevinovat), „psihologul trebuie să se supună unor reguli specifice de probare, în stabilirea faptului că un anumit test are o însemnătate specială pentru un scop specific” (Aiken, op. cit., p. 133). Deci validitatea ar putea fi definită mai corect ca o extindere progresivă, pe măsură ce testul primeşte noi utilizări, a cunoaşterii noastre în legătură cu ceea ce măsoară el de fapt.

6.1. Validitatea de faţadă

Este mai mult o aparenţă de validitate decât o formă de validitate în sens strict, deoarece ea indică *ce pare* a măsura o probă, şi nu ceea ce măsoară ea efectiv. Uneori psihologii raportează exact ceea ce observă, fără a interpreta sau generaliza, şi atunci intervine tot validitatea de faţadă.

Chiar dacă tehnic ea este slab definită (şi de aceea uneori nici nu este analizată de mulţi autori), validitatea de faţadă pare a avea o importantă funcţie practică, deoarece motivează subiecţii să accepte mai uşor un test nou, sau permite extensia utilizării instrumentelor spre alte zone decât cele care le-au consacrat: „adesea spunem că un test are

validitate de faţadă dacă itemii lui sunt relaţionaţi rezonabil cu scopurile percepute ale testului" (Kaplan şi Saccuzzo, op. cit., p. 135). Uneori acest tip de validitate poate da o mai bună valoare de piaţă unui test.

6.2. Validitatea relativă la conţinut

Aşa cum reprezentativitatea eşantionului de populaţie ales permite generalizări asupra populaţiei–ţintă pe care o reprezintă, itemii testului – prin maniera lor de construcţie şi de selecţie – pot acoperi într-o mai mare sau mai mică măsură domeniul sau universul trăsăturii sau al aptitudinii măsurate. În felul acesta răspunsurile la un eşantion de itemi dintr-un test cu validitate de conţinut sunt reprezentative pentru răspunsurile pe care subiectul le-ar fi dat dacă întreg universul trăsăturii ar fi fost măsurat. Acest tip de validitate este decisiv în testele educaţionale şi de achiziţii, ridicând probleme specifice evaluatorului. Se va genera deci o bancă sau set de itemi care să acopere bine întregul domeniu investigat (cunoştinţele la un obiect şcolar, gradul de elaborare a unor deprinderi, nivelul de cristalizare a unor atitudini), din care vor fi selecţionaţi şi agregaţi într-un test doar un număr limitat de itemi, reprezentativi pentru întregul domeniu.

Unele date de personalitate, obţinute prin chestionare (de opinii, atitudini, interese, motivaţii sau valori), beneficiază mai mult de acest tip de validitate decât celelalte. În testele de achiziţii itemii sunt creaţi ţinând cont simultan de obiectivele, finalităţile procesului instructiv, dar şi de mecanismele proceselor cognitive la care se face apel în elaborarea răspunsului la itemi. Aceasta presupune încă din start prezenţa experţilor care vor judeca şi hotărâ care itemi vor fi incluşi în test. De aceea putem spune că dintre formele fundamentale de validitate, cea de conţinut este singura care are o susţinere mai degrabă logică decât statistică.

Pentru a da validitate de conţinut unui test, constructorul se angajează într-un proces de durată, ce presupune o foarte bună cunoaştere a domeniului, raţionament logic, intuiţie şi perseverenţă, căci itemii trebuie continuu revizuiţi. Pentru a face generalizări întemeiate plecând de la scorurile la test, constructorul trebuie să cunoască bine toţi factorii care ar putea afecta performanţa subiecţilor.

6.3. Validitatea relativă la criteriu

În acord cu această faţetă a validităţii, un test este valid relativ la criteriu dacă pe baza lui se pot lua decizii corecte sau se pot face predicţii/ prognoze asupra persoanelor examinate. Un test este valid dacă poate înlocui variabila criteriu în virtutea unor corelaţii mari între test

şi aceasta, astfel încât cunoscând scorul la test să deducem valoarea variabilei criteriu pentru persoana examinată. După maniera în care sunt colectate datele la test şi cele la criteriu distingem două situaţii, care produc două subcategorii ale validităţii relative la criteriu:

- Când scorurile la test şi la criteriu se obţin simultan, este vorba de *validitate concurentă*.
- Când între obţinerea scorurilor la test şi cele la criteriu se interpune o perioadă de timp, vorbim de *validitate predictivă*.

Fiecare dintre aceste tipuri presupune ***strategii de validare*** diferite[3], care vor fi detaliat analizate în capitolul următor. De exemplu, un test de anxietate (Cattell) are validitate concurentă dacă scorurile la testul psihologic îşi găsesc confirmarea, în aceeaşi perioadă de timp, prin diagnosticul psihiatric. Atunci când criteriul se „maturează" după un timp (luni sau ani) şi scorurile lui nu pot fi determinate simultan cu cele la test, folosim o a doua strategie de validare, cea predictivă. În selecţia profesională, de exemplu, se va calcula coeficientul de corelaţie liniară dintre scorurile la testul aplicat iniţial şi măsuri (exprimate prin note sau calificative) ale performanţelor obţinute de aceleaşi persoane în activitatea sau profesia pentru care au fost selecţionate. Deoarece o funcţie fundamentală a testului este aceea de a face diagnoza în scopul unei prognoze (deci faptul de prezice), *validitatea predictivă* capătă o importanţă specială în acest context.

Întrucât măsurile la test şi cele la criteriu intră în corelaţie, pentru a determina cu exactitate limitele predicţiei, aceasta impune ca ambele elemente corelate să satisfacă standarde înalte de fidelitate şi de validitate. Cu alte cuvinte, o măsură criteriu este pertinentă dacă este şi ea fidelă şi validă, fapt ce introduce problema validităţii într-o relaţie circulară cu criteriul ales. Aceasta înseamnă că factorii de eroare care afectează criteriul vor fi controlaţi în aceeaşi manieră ca şi predictorii (testele). Atunci când controlul lor nu dă rezultatele scontate, putem spori fidelitatea criteriului fie luând mai multe eşantioane de măsurători ale acestuia (măsuri mai dese), fie adăugând noi tipuri de măsuri, incluse într-un criteriu compozit, pentru a spori fidelitatea lui în aceeaşi manieră în care procedăm cu un test căruia îi adăugăm itemi suplimentari.

O caracteristică importantă a criteriului este *absenţa contaminării*, adică necunoaşterea rezultatelor la test de către cel care evaluează măsurile–criteriu. Un psiholog va obţine, independent de diagnosticul medicului psihiatru, scorurile la testul său de anxietate, deci va face o analiză oarbă. Prin efectul de halo, indulgenţă, predicţia pentru sine

[3] Vezi capitolul următor.

sau profeţia care se autoîmplineşte, rezultatele prezise pot fi produse într-o oarecare măsură de predictorul însuşi, caz în care comparaţia nu mai este validă. Havârneanu apreciază că „cele mai bune măsuri–criteriu sunt cele care prezintă cât mai multe avantaje practice, care sunt mai simplu de folosit, disponibile şi mai puţin costisitoare" (op. cit., p. 113). Dintre acestea enumerăm:

- măsurarea directă a producţiei sau a randamentului;
- evaluările performanţelor angajaţilor făcute de către şefii lor sau de către serviciile de personal ale întreprinderilor;
- măsura–criteriu poate aparţine unui grup, în sensul că testul care reuşeşte să clasifice persoanele în grupuri bine precizate, sunt valide.

Strategiile de validare predictivă cele mai utilizate calculează fie un indice de eficacitate, fie scoruri de separare a categoriilor, indici de separare a grupurilor sau indicele de utilitate (vezi Havârneanu, op. cit., pp. 113-126).

Dintre factorii care afectează validitatea relativă la criteriu, Aiken indică diferenţele de omogenitate ale grupurilor (validitatea tinde să fie cu atât mai mică cu cât grupurile sunt mai eterogene), lungimea testului (testele mai lungi sunt mai valide) şi contaminarea criteriului, de care am vorbit deja. Creşterea validităţii unui test, în calitatea lui de instrument diagnostic sau prognostic, presupune costuri ce trebuie atent evaluate, căci uneori metode mai puţin costisitoare (observaţia, interviul sau inventarul biografic) pot aduce plusul de informaţie necesar în atingerea scopurilor propuse.

6.4. Validitatea relativă la construct

Multe dintre testele de personalitate măsoară dimensiuni ale acesteia (anxietate, depresie, nevrotism, motivaţie, introversie-extraversie), mai slab definite iniţial, dar din ce în ce mai bine circumscrise, pe măsură ce teoriile în domeniu au avansat. Conceptualizările progresive ale acestora (dar şi ale inteligenţei, memoriei, creativităţii etc.) au dus la apariţia de constructe tot mai evoluate.

Unele teste sunt construite plecând de la date empirice (ca MMPI sau *Inventarul de Personalitate California* al lui Gough), altele se originează în teorii anterior elaborate oricărei cercetări empirice (*Chestionarele de personalitate Cattell* sau *Indicatorul Tipologic Myers-Briggs*). Validitatea relativă la construct verifică, pe de o parte, dacă testul se referă realmente la constructul pe care vrea să-l măsoare, şi apoi dacă scorurile subiecţilor testaţi reflectă

corect mărimea acestui construct la persoanele în cauză. Strategia determinării acestui tip de validitate impune un algoritm (vezi Albu, 1998, pp. 165-166):

- Descrierea amănunţită a *domeniului constructului* respectiv (adică a tuturor atributelor ce au legătură cu acesta şi specificarea relaţiilor dintre ele), printr-o reţea nomologică, pe care un grup de experţi o va evalua pentru a determina nivelul la care ea acoperă constructul. Apoi se realizează un *model al rezolvării testului*, cu mecanismele teoretice implicate, pentru a vedea legătura acestuia cu constructul, dar şi cu răspunsurile posibile la test. Este posibil să se constate o bună reprezentare a constructului în test, dar şi diverse grade de subreprezentare (mai ales când sarcinile sunt prea grele sau prea uşoare).
- Se analizează *consecvenţa răspunsurilor la itemii* testului, adică dacă procesele implicate în rezolvarea acestora acţionează similar la toţi itemii, pe parcursul întregului test.
- Se urmăreşte dacă structura globală a scorului la test reflectă structura domeniului constructului măsurat. Pentru aceasta se apelează la analiza factorială (de unde şi numele de *validare factorială*), care determină câteva variabile latente numite factori, alcătuiţi din itemi ce dau scoruri care corelează între ele, ce acoperă varianţa întregului test. Se caută semnificaţia acestor variabile, reţinându-se cele care au o mai mare legătură cu constructul, asigurându-se o corespondenţă între importanţa avută în definirea constructului respectiv şi ponderea lor ca itemi în testul final.
- *Se extind generalizările* acestor prime etape, schimbând fie populaţiile examinate, fie examinatorul sau contextele de examinare, pentru a vedea dacă relaţiile sesizate rămân constante.
    - În final se caută stabilirea relaţiilor dintre testul cercetat şi alte tipuri de măsurători sau de observaţii. Testul poate avea o *validitate convergentă* (între scorurile sale şi alte tipuri de teste ce măsoară acelaşi construct există o relaţie liniară), sau *discriminantă* (când scorurile la teste nu corelează cu teste care se ştie că nu măsoară constructul respectiv).

CAPITOLUL 7

ALTE FEŢE ALE VALIDITĂŢII

Orice student la psihologie ştie că meritele unui test psihologic încep să fie judecate prin fidelitate dar sfârşesc prin aprecierea validităţii sale. Aceasta şi datorită faptului că, în timp ce fidelitatea se judecă simplu, prin sumarizarea unor coeficienţi de corelaţie, validitatea presupune demersuri interpretative mai complicate şi de mai lungă durată. Însuşi conceptul de validitate este unul evolutiv, aflat în plină dezvoltare şi „de aceea stârneşte mult mai multă controversă decât mai stabilul şi mai bine aşezatul său văr, fidelitatea" (Cronbach, apud Gregory, op. cit., p. 106). Acesta este motivul pentru care câteva aspecte legate de validitate, abia schiţate în capitolul anterior, sunt abordate mai pe larg în capitolul de faţă.

7.1. Validitatea relativă la construct (conceptuală)

Este cu siguranţă tipul de validitate care necesită cea mai mare cantitate de efort în vederea validării unui instrument psihometric. Deşi se admite faptul că soarta unei ştiinţe este dependentă de nivelul cuantificărilor sale, adică de capacitatea sa de a măsura adecvat ceea ce intră în domeniul obiectului său de studiu, şi reciproca este valabilă. Astfel, nu am putea spune nimic concret despre variabile complexe ca inteligenţa, memoria, stilurile de evaluare sau cele cognitive dacă nu am avea şi posibilitatea să le testăm, adică să le supunem măsurării. Dar, pe de altă parte, măsurarea lor este departe de a fi posibilă fără o cunoaştere teoretică prealabilă a acestor constructe.

Raportul dintre teorie şi măsurare este cel mai bine surprins prin conceptul de **validitate de construct** sau conceptuală, care este chiar inima operaţionalizării variabilelor. A operaţionaliza înseamnă de fapt a aduce un construct teoretic în situaţia de putea fi surprins şi evidenţiat prin măsurare. În cazul în care operaţionalizarea unei teorii bune a fost făcută corect, printr-un instrument corect, capabil să o surprindă, atunci putem formula predicţii sub forma ipotezelor ce rezultă din teoria respectivă. Dacă ipotezele se verifică, atunci şi teoria, dar şi operaţionalizarea ei sub forma instrumentului de măsură respectiv sunt în regulă.

Dacă ipotezele nu se verifică, sunt posibile două explicaţii alternative:

- instrumentul este bun, dar deoarece teoria nu se verifică, aceasta trebuie schimbată;

- teoria este una valabilă, dar instrumentul nu o operaţionalizează corespunzător şi atunci nu putem dovedi ce ne-am propus şi deci trebuie schimbat chiar instrumentul de măsură.

De exemplu, cercetările lui Witkin şi ale colaboratorilor săi au evidenţiat mai multă diferenţiere şi mai multă independenţă de câmp pentru genul masculin. Studiile neuropsihologice actuale despre lateralizarea emisferelor cerebrale indică de asemenea o specializare emisferică mai accentuată pentru genul masculin, ca şi un acces bicerebral la limbă pentru genul feminin, de unde superioritatea bărbaţilor pentru sarcinile spaţiale şi a femeilor pentru cele legate de limbă şi comunicare. Faptul că la un test spaţial ca şi Cuburile Kohs, de exemplu, băieţii obţin performanţe superioare comparativ cu fetele, în timp ce acestea au performanţe superioare la un test de inteligenţă verbală, este consistent cu teoriile anterior menţionate.

În alte situaţii se aşteaptă ca itemii unui test să se organizeze într-o ordine progresivă, ceea ce ar corespunde unei structuri ierarhice, rezultată din chiar dezvoltarea funcţiei sau abilităţii respective. De exemplu, dacă vom operaţionaliza teoria psihogenetică piagetiană într-o scală de itemi, aceasta va fi una ierarhică, pentru că orice stadiu îl înglobează pe cel precedent, facându-l posibil pe cel care urmează. Validitatea conceptuală va fi serios pusă la îndoială în cazul în care un subiect pasează itemii pentru stadiul operaţiilor formale şi nu-i trece pe cei din stadiul operaţiilor concrete, fapt care ar contrazice chiar ideea de dezvoltare stadială şi ierarhică. În mod sporadic s-ar putea întâmpla să existe şi itemi de acest fel, care vor constitui însă erori.

Se poate deci afirma că rezultatele la acest tip de test sunt *reproductibile*, adică – pornind de la scorul total la test – putem să prezicem care itemi vor fi pasaţi şi care vor fi eşuaţi. Pentru aceasta însă trebuie să ştim ordinea de dificultate a itemilor, iar această ordine să fie aceeaşi la toţi subiecţii. Cazurile (cât mai puţine!) în care ierarhia itemilor este încălcată produc erori de reproductibilitate, căci ele limitează posibilitatea predicţiei corecte a itemilor care vor fi pasaţi şi al celor ce vor fi eşuaţi. Guttman (1950) a propus ca rezultatele la un test să fie considerate ierarhice atunci când mai puţin de 10% din rezultate nu sunt reproductibile şi a propus o formulă de calculul al unui coeficient de reproductibilitate:

$$\mathrm{CR} = 1 - \frac{N_e}{N_j N_p}$$

în care:

CR = coeficientul de repreoductibilitate;

N_e = numărul erorilor de reproductibilitate (numărul de itemi care încalcă ierarhia, adică furnizează rezultate în dezacord cu scorul total obţinut şi cu ordinea de dificultate a ansamblului itemilor);

N_j = numărul de itemi;

N_p = numărul de persoane (subiecţi).

Aceasta poate fi considerată o bună metodă de validitate conceptuală pentru probele piagetiene, condiţia recomandată de Guttman fiind un CR de peste 0,90, deoarece peste această valoare invarianţa ordinii de reuşită se consideră a fi bine demonstrată.

Alte metode de demonstare a validităţii conceptuale o constituie metoda corelaţiilor simple, metoda multitrăsătură-multimetodă şi studiul trăsăturilor latente.

7.**1.1. Metodele corelaţionale simple**

Acestea permit evidenţierea faptului dacă un test măsoară acelaşi lucru cu un alt test despre care se ştie deja că este o măsură validă a constructului respectiv. Acest tip de validitate se numeşte **validitate convergentă**. Când însă testul respectiv este pus în corelaţie cu un alt test care măsoară o aptitudine complet diferită, această corelaţie trebuie să fie foarte mică şi atunci vom pune în evidenţă **validitatea divergentă** (sau discriminantă). De exemplu, este cu totul de aşteptat ca un test de inteligenţă să coreleze puternic cu randamentul şcolar la 12 ani, aşa cum menţionează şi literatura de specialitate, căci cele două variabile sunt învecinate. Când însă testul nostru corelează puternic cu motivaţia, cu locus of control sau cu alte variabile îndepărtate apar mari suspiciuni în legătură cu ce măsoară de fapt acest test.

7.**1.2. Matricile multitrăsătură-multimetodă**[4]

Procedeul descris aici a fost inventat încă din 1959 de Campbell şi Fiske, el fiind în acelaşi timp şi o metodologie importantă pentru înţelegerea procesului de validare. Acesta este un design ce ia simultan în considerare validitatea convergentă şi pe cea divergentă (discriminantă). Se cheamă multidimensională pentru că ia în evaluare două sau mai multe trasături, prin două sau mai multe metode.

În exemplul dat de Gregory în figura de mai jos, trei trăsături A, B şi C (Interesele, Creativitatea şi Dominanţa) sunt măsurate cu metodele 1, 2 şi 3 (inventar autoaplicat, evaluarea covârstnicilor şi un test proiectiv). Deoarece fiecare dintre cele 3 trăsături sunt măsurate prin toate cele 3 metode, rezultă de fapt 9 teste, exprimate prin coeficienţii de

[4] *Multitrait-Multimethod Matrix* în limba engleză.

corelaţie din triunghiurile cu linie continuă de pe diagonala mare. Când fiecare dintre aceste teste este administrat a doua oară aceluiaşi grup de subiecţi şi scorurile fiecărei perechi de teste rezultate prin test-retest sunt corelate, rezultă matricea multitrăsătură-multimetodă care devine o sursă extrem de bogată de informaţii, atât pentru fidelitate, cât şi pentru validitatea convergentă şi divergentă.

Metoda	**Trăsăturile**	***Metoda 1***			***Metoda 2***			***Metoda 3***		
		A1	B1	C1	A2	B2	C2	A3	B3	.C3
Metoda 1	A1	(.89)								
	B1	.51	(.89)							
	C1	.38	.37	(.76)						
Metoda 2	A2	**.57**	.22	.09	(.93)					
	B2	.22	**.57**	.10	.68	(.94)				
	C2	.11	.11	**.46**	.59	.58	(.84)			
Metoda 3	A3	**.56**	.22	.11	**.67**	.42	.33	(.94)		
	B3	.23	**.58**	.12	.43	**.66**	.34	**.**67	(.92)	
	C3	.11	.11	**.45**	.34	.32	**.58**	.58	.60	(.85)

Notă: - literele A, B şi C se referă la trăsături iar cifrele 1, 2, şi 3 subscrise se referă la metode.
- matricile constă din coeficienţi de corelaţie.

Sursă: Gregory, R.J. (1996). ***Psychological Testing. History, Principles, and Applications***. Boston, London, Toronto, Sidney, Tokio, Singapore: Allyn and Bacon, p. 122.

Figura 7.1. *Modul de raportare a validităţii prin Matricea multimetodă-multitrăsătură*.

Orientarea în tabelul de mai sus se face astfel:

- coeficienţii dintre paranteze de pe diagonala mare sunt coeficienţii de *fidelitate* pentru fiecare test în parte: cu cât este mai mare, cu atât fidelitatea este mai bună, deşi în realitate ne mulţumim cu r cuprinşi între 0,80 şi 0,90;
- coeficienţii boldaţi de pe cele trei diagonale mici reprezintă dovezi ale validităţii *convergente*: aceeaşi trăsătură fiind măsurată prin diferite metode, corelaţia trebuie să fie una pozitivă şi puternică;
- în triunghiurile cu linie continuă sunt incluse corelaţiile dintre diferitele trăsături măsurate cu aceeaşi metodă: acestea trebuie să aibă valori medii sau ridicate;
- cele 6 triunghiuri din linii punctate includ corelaţiile dintre diferitele trăsături măsurate cu diferite metode; acestea trebuind să fie cât mai mici în fiecare triunghi, deoarece ele sunt dovezi importante despre validitatea discriminantă (*divergentă*).

Această metodologie este o contribuţie importantă la înţelegerea procesului de validare, dar reclamă un efort de cercetare uriaş. De aceea cei mai mulţi creatori de teste preferă să culeagă informaţii separate, „pe bucăţi" despre fidelitate şi validitate, pe care să le reunească ulterior într-un tabel compozit, decât o matrice care să le integreze din capul locului pe toate deodată simultan.

7.1.3. Studiul trăsăturilor latente

Ne aşteptăm ca itemii care măsoară un acelaşi construct sau caracteristică să fie puternic corelaţi între ei, ceea ce ar putea primi o reprezentare grafică de tip *scatter,* care evidenţiază norul de puncte dintre o abscisă şi o ordonată reprezentând cele două variabile intercorelate. Analiza factorială este o tehnică specială care permite să se meargă mult mai departe decât simpla inspecţie vizuală a matricilor de intercorelaţii sau a norului de puncte. Scopul metodei este acela de a determina numărul minim de determinanţi numiţi *factori,* dar şi *trăsături latente,* necesari explicării intercorelaţiilor din matricile respective.

Demersul tipic al analizei factoriale constă din a administra o baterie de teste câtorva sute de subiecţi şi apoi de a calcula intercorelaţiile dintre toate perechile posibile de teste, ceea ce va produce o matrice de intercorelaţii. Deoarece fiecare test *nu* este o trăsătură independentă (caz în care intercorelaţiile din matrice ar tinde spre zero) şi deoarece mai multe teste măsoară în diverse grade o aceeaşi trăsătură, formând *clusteri* (ciorchini), analiza factorială tinde să descopere un număr mai mic de factori decât numărul testelor intercorelate şi apoi să producă un tabel care să indice încărcătura (saturaţia) fiecărei variabile în factorul respectiv. Aceasta poate varia, ca orice coeficient de corelaţie, între -1,00 şi +1,00. Rezultatul cel mai important al analizei factoriale este un tabel cu factorii şi saturaţia fiecărei variabile în factorii respectivi, ceea ce ne ajută să descriem structura factorială şi să producem astfel dovezi despre validitatea de construct a unei baterii de teste sau de variabile reunite.

Creatorul analizei factoriale, Charles Spearman (1907) a dezvoltat o metodă capabilă să pună în evidenţă existenţa unui factor general al inteligenţei, prezent în orice tip de sarcină mintală (şi de aceea numit şi factor **g**), în combinaţii diferite cu un factor specific numit **s**. Procedeul dezvoltat de el, numită *metoda componentelor principale*, i-a permis să producă dovezi în favoarea ipotezei sale, numită modelul bifactorial al inteligenţei.

În 1928, un creator american în domeniul analizei factoriale, Thurstone, a utilizat aceeaşi metodă ortogonală (două axe aflate în unghi drept, adică independente între ele) dar cu un mic amendament: în loc ca o axă să explice varianţa factorului major şi cealaltă varianţa reziduală (varianţa rămasă neexplicată de factorul major) el a propus rotirea celor

două axe ortogonale până la obţinerea structurii celei mai simple, ameliorând astfel gradul de adaptare al datelor la structura factorială. Metoda se mai numeşte şi *Varimax* deoarece ea caută ca pe fiecare dintre cele două axe ortogonale varianţa explicată să fie una maximală. Aceasta face ca fiecare factor să explice un grup de rezultate şi doar atât, căci cu această metodă nu mai poate fi identificat factorul general, ceea ce pune în discuţie modelul propus de Spearman pentru a defini inteligenţa umană.

Din punct de vedere matematic ambele metode sunt valide, dar fiecare are propriile posibilităţi şi limite: una ajută la identificarea factorului general pe care cealaltă îl neagă, dar cealaltă ajută la mai buna identificare, înţelegere şi numire a factorilor rezultaţi. De aceea utilizarea lor este corelativă. Analiza factorială permite doar verificarea faptului dacă datele sunt consistente cu structura factorială postulată. Când datele sunt compatibile cu mai multe structuri latente, analiza factorială nu ne poate ajuta să alegem una singură, această alegere fundamentându-se pe baza unei teorii prealabil elaborată. Într-un demers de validare, analiza factorială ne va aduce deci informaţiile necesare, dar nu şi suficiente luării unei decizii.

Analiza factorială se sprijină pe două postulate de bază:

- *postulatul cauzalităţii factoriale*: variabilele observate sunt combinaţii lineare de variabile cauzale subiacente;
- *postulatul parcimoniei*: dintre două soluţii factoriale trebuie să o alegem pe cea mai simplă. Deşi larg acceptată de majoritatea cercetătorilor, nu este posibil de demonstrat că acest postulat este bine fundamentat şi nici nu este valabil faptul că întotdeauna structura cea mai simplă să fie mai plauzibilă decât cea mai complexă.

Oricum, în funcţie de tipul de analiză ales ajungem la soluţii factoriale diferite, iar soluţia finală pe care o alegem nu este una de natură matematică, ci de natură teoretică. În plus, analiza factorială pune nu numai probleme de interpretere, ci şi câteva importante probleme metodologice, rezumate mai jos.

- ***Mărimea eşantionului***: cu cât eşantionul pe care se face analiza factorială este mai mic, cu atât mai mult coeficienţii de corelaţie din matrice pierd din semnificaţia statistică, soluţiile factoriale obţinute trebuind examinate cu mai multă prudenţă. Deşi nu poate fi indicată o mărime ideală, cercetătorii au căzut de acord cu raportul de 5 la 1 între numărul subiecţilor din eşantion şi numărul variabilelor supuse factorizării. Dar şi acest lucru trebuie luat cu rezervă, pentru că corelaţiile mici (sub 0,30) pun în discuţie nu mărimea eşentionului, ci chiar necesitatea desfăşurării unei analize factoriale, după cum corelaţiile mai mari de 0,50 se pot desfăşura cu rezultate bune şi pe eşantioane mai mici decât cele

ce ar rezulta din raportul 5/1. Deci, înainte de a efectua o analiză factorială, obligatoriu vom inspecta cu atenţie matricea de intercorelaţii a variabilelor factorizate.

- ***Normalitatea multivariată***: aceasta semnifică faptul că toate variabilele şi toate combinaţiile lor se distribuie normal. Cum nu putem testa toate combinaţiile de variabile, ceea ce ne rămâne de făcut este faptul de a aprecia normalitatea fiecărei variabile în parte din coeficientul de asimetrie şi de boltire (*skewness* şi *kurtosis*).
- ***Liniaritatea***: relaţia dintre variabile trebuie să fie liniară, în caz de curbiliniaritate rezultatele analizei factoriale fiind de pus la îndoială. Liniaritatea dintre variabile se poate verifica prin *scatter*-ul cu norul de puncte pentru fiecare combinaţie de câte 2 variabile.

7.2. Validitatea diferenţială

Validitatea unui test este determinată pe ansamblul populaţiei pe care el a fost aplicat. Aceasta ar putea să însemne că că validitatea astfel determinată este una valabilă pentru toţi subiecţii acestei populaţii, postulat care afost tot mai mult pus la îndoială după anii 70. În felul acesta s-a ajuns la concluzia că validitatea unui test poate varia înăuntrul aceleiaşi populaţii prin faptul că unii subiecţi pot aprţine unor subgrupuri speciale. De exeemplu, validitatea determinată pentru un test de lectură poate fi diferită pentru stângaci, ca subpopulaţie a grupului total. Validitatea pentru un test de matematică poate fi afectată în cazul unui grup special, care obţine performanţe sistematic mai scăzute, nu din cauza capacităţii de raţionament mai scăzut a membrilor acestui subgrup, ci ca urmare a faptului că aceştia au o tulburare specifică de citit (dislexie) şi înţeleg cu mai mare dificultatea enunţurile itemilor testului.

Aceste constatări conduc la lărgirea cadrului conceptual al validităţii nu numai la diferitele utilizări care se dau unui test, ci şi pentru diferitele subgrupuri populaţionale pe care se aplică, în raport cu care el este mai mult sau mai puţin valid. Pentru acest tip de validitate termenul care s-a încetăţenit este acela de *validitate diferenţială* iar populaţiile care se abat de la regula generală se cheamă că sunt *biasate* (distorsionate).

Studiul acestei forme de validitate este şi el foarte complex şi rămâne – ca orice altă formă de validitate – un demers doar parţial împlinit, adică neterminat. Pentru orice utilizare a unui test pe un grup nou trebuie să se aducă dovezi că el nu este unul biasat. Simpla constatare a diferenţei dintre două grupuri nu însemană neapărat că unul dintre ele este neapărat biasat. Performanţele mai slabe la un test de vocabular ar putea avea ca explicaţie şi oportunităţile de învăţare diferite pentru cele două grupuri.

Testele ce ar putea pune în evidenţă diferenţele dintre cele două grupuri pot avea şi o utilitate socială, pentru că pot pune în mişcare acţiuni de remediere pentru a da membrilor fiecărui grup şanse egale. Ca şi validitatea în general, cea diferenţială se determină tot prin trei procedee, ce au şi acelaşi nume: validitatea de conţinut, validitatea relativă la criteriu şi validitatea relativă la construct (sau conceptuală). Pentru a nu îngusta prea mult problematica validităţii diferenţiale, nu vom intra în detalii cu prezentarea sa.

7.3. Teoria deciziei aplicată la testele psihologice, asociată validităţii predictive

Mulţi psihologi susţin că scopul testării psihologice nu este măsurarea în sine, ci măsurarea în serviciul luării unei decizii optime. Astfel, managerul vrea să ştie pe cine să angajeze, examinatorul pe cine să admită iar psihiatrul pe cine să interneze. Testele psihologice joacă în mod frecvent un rol important în aceste tipuri de decizii instituţionale. Deşi în luarea unei decizii adesea sunt luate în considerare mai multe puncte de separaţie a grupurilor, numite şi puncte de tăietură (*cutoff*), pentru a simplica lucrurile noi vom pleca de la exemplul unui singur test. În realitate, scorurile la un test se dispun pe ceea ce în statistică numim amplitudinea împrăştierii (R = *range*). Cu toate acestea este posibil să identificăm punctul care desparte jumătatea căreia i se prezice succes de cea căreia i se prezice nereuşită.

Să presupunem de asemena că cei cărora li s-a prezis succes au fost şi selectaţi sau angajaţi într-o oarecare proporţie, ce poate varia de la 0 la 1, după numărul de persoane considerate a fi susceptibile să reuşească în funcţie de criteriul de măsură adoptat. Dacă procedăm foarte reducţionist, succesul poate fi împărţit într-o variabilă dihotomică – „predicţie succes" versus „predicţie eşec" – în timp ce criteriul prin care se va determina succesul ulterior poate fi împărţit la rândul său în „a reuşit" şi „a avut eşec". Aşadar, orice studiu de validitate predictivă poate produce o matrice cu două intrări de tipul celei de mai jos.

		Performanţa la criteriu	
		A avut succes	**A avut eşec**
Predicţia la test	**Va avea succes**	Predicţie corectă	Falşi pozitivi
	Va avea eşec	Falşi negativi	Predicţie corectă

Figura 7.2. *Rezultatele posibile ale unui test de selecţie utilizat pentru a preciza un criteriu-măsură.*

Dacă un test a avut o bună validitate predictivă, atunci diagonala stânga-sus – dreapta-jos se maximizează: cele mai multe persoane cărora li s-a prezis succes au şi avut succes iar cele mai multe persoane cărora li s-a prezis eşec au şi eşuat, deci ambele predicţii s-au adeverit. Dar cum nici un test de selecţie nu este un predictor perfect, sunt posibile şi celelalte două situaţii: s-a prezis succes unor persoane care au eşuat (*falşii pozitivi*) ori s-a prezis eşec şi acesta nu a avut loc (*falşii negativi*). Ambele categorii de falşi, pozitivi şi negativi, sunt considerate în egală măsură eşecuri ale predicţiei, deoarece ambele au făcut predicţii inacurate. Se poate astfel calcula următoarea proporţie:

Rata de reuşită = reuşite / (reuşite + eşecuri)

În lumea reală fals-pozitivii şi fals-negativii sunt evitaţi, dar nu până la eliminare, pentru că aceasta ar presupune un test care are o corelaţie perfectă (r = + 1,00) cu criteriul prezis. Cum acest lucru nu este posibil, ceea ce rămâne de făcut pentru a minimiza erorile de decizie este dezvoltarea de teste cu o mare valoare predictivă.

În luarea deciziilor, adepţii teoriei deciziei se sprijină pe două asumpţiuni:

- Dezvoltarea unei *scale de utilitate*, astfel încât indicatori importanţi ai instituţiei să poată fi exprimaţi în profituri sau pierderi. De exemplu, dacă după un chestionar de tipul EPI al lui Eysenck, prin care se poate determina şi introversia/extraversia, o companie poate identifica pierderile asociate cu vânzătorul de maşini introvert, comparativ cu câştigurile aduse de vânzătorul extravert. Astfel, falşii pozitivi sunt cei ce aduc pierderi instituţiei (vând mai puţine maşini decât ar fi necesar pentru plata propriului salariu), ca şi falşii negativi (cei care, dacă ar fi fost încadraţi, ar fi contribuit la creşterea profiturilor companiei). Acesta este un exemplu de operaţionalizare a testului şi a selecţiei pe o scală de utilitate.
- În deciziile de selecţie instituţională strategia cea mai satisfăcătoare este aceea de a *maximiza câştigul* scalei de utilitate şi de a *minimiza pierderile*. Ea trebuie să răspundă la întrebarea „la ce proporţie de selecţie se produce cel mai mare câştig mediu al scalei?”

Tabelele Taylor-Russell, publicate prima dată de aceşti autori americani în 1939, permit unui utilizator de teste să determine proporţia expectată pentru aplicanţii selectaţi dacă se specifică trei elemente: 1. validitatea predictivă a testului, 2. rata (proporţia) selecţiilor şi 3. proporţia bazală pentru aplicanţii de succes.

Ce sunt aceste elemente?

- *validitatea predictivă* este determinată anterior prin studii şi se exprimă printr-o corelaţie de tipul r_{xy};
- *proporţia selecţionaţilor* este cea rezultată din numărul de aleşi supra numărul de aplicanţi, înmulţit cu 100;
- *rata de bază* este proporţia celor ce ar fi fost selecţionaţi prin metodele curente, fără beneficiul adus de noul test. Cum acest lucru este mai greu de determinat, la extrem se foloseşte pentru rata de bază proporţia aplicanţilor de succes, dacă ei ar fi fost aleşi aleator.

Cunoaşterea celor trei factori de mai sus face posibilă utilizarea unor tabele special construite prin care se determină cu cât îmbunătăţeşte noul test selecţia, comparativ cu metoda anterioară. Pe lângă faptul că leagă testul psihologic de aspecte economice foarte pragmatice, determinând efectiv valoarea practică a unui test, tabelele Taylor-Russell evidenţiază şi aspecte care intrigă, cum ar fi creşteri substanţiale ale acurateţei selecţiei, chiar şi la validităţi foarte mici, în anumite condiţii.

7.4. Precauţii legate de evaluarea coeficienţilor de validitate predictivă

Standards for Educational and Psychological Testing din SUA atrag atenţia asupra câtorva probleme importante şi precauţii legate de interpretarea coeficienţilor de validitate:

- *Cercetaţi schimbările apărute în sensul cauzalităţii test-criteriu*: de regulă presupunem implicit că relaţia dintre test şi criteriu este încă în funcţiune atunci când aplicăm noi testul. De exemplu, testul s-a dovedit util în selecţia de supraveghetori femini, ceea ce s-a soldat cu succes într-o primă fază, când grupul era demixtat (numai femei), dar nu ia în consideraţie situaţia actuală în care grupul de muncă este unul mixt.
- *Ce însemană criteriul ales*? Coeficientul de validitate contra criteriului obţinut poate să nu însemne nimic dacă criteriul nu este el însuşi fidel şi valid. A corela un test cu un altul despre care nu ştim prea multe lucruri sub aspectul validităţii nu ne ajută prea mult în determinarea validităţii convergente a primului test.
- Luaţi întotdeauna în consideraţie *populaţia pe care s-a obţinut o anume validitate*: o validitate ridicată pe o populaţie de albi se păstrează ea oare şi pe o populaţie de negri?
- Asiguraţi-vă că *mărimea eşantionului populaţional este suficientă şi adecvată*: cu cât este mai mare grupul iniţial pe care s-a studiat validitatea, cu atât mai bună este validitatea încrucişată.

- Niciodată să *nu confundaţi criteriul cu predictorul*.
- Verificaţi *spectrul de valabilitate* atât pentru predictor, cât şi pentru criteriu: deoarece corelaţia depinde de mărimea variabilităţii, ea este subestimată sever atunci când predictorul sau criteriul au un spectru de variabilitate foarte îngust.
- *Generalizaţi cu prudenţă* o validitate găsită*:* cu alte cuvinte trebuie demonstrat că validitatea descoperită într-un studiu nu este dependentă de situaţie.
- Luaţi în calcul şi *validitatea diferenţială*: relaţiile descoperite pentru o populaţie demografică s-ar putea să nu fie deloc aceleaşi pentru toate subgrupurile care o compun.

7.5. Interpretarea coeficienţilor de validitate

Pentru că strategiile de validare a unui test sunt atât de diverse, nu există o interpretare uniformă a acestora. În practică întâlnim destul de rar coeficienţi mai mari de 0,60, cei mai mulţi (consideraţi satisfăcători sau chiar mari) oscilând între 0,30 – 0,40.

Un coeficient de validitate este semnificativ statistic când şansele de a-l obţine din întâmplare sunt sub 5 procente. Pentru a fi mai specifici, vom arăta că pătratul unui coeficient de validitate (coeficientul de determinare) indică procentajul din variabilitatea criteriului pe care aşteptăm să o ştim în avans prin cunoaşterea scorurilor la test. La o corelaţie de 0,60 dintre o baterie de teste şi media şcolară anuală, vom şti că $0{,}60^2 = 0{,}36$, adică 36% din variabilitatea mediei şcolare depinde de inteligenţă, aşa cum am măsurat-o cu bateria noastră, restul depinzând de alte cauze, printre care şi de factorii de eroare.

Uneori o validitate de 0,30 ar putea înseamna economii de milioane de dolari în orientarea şi selecţia profesională, alteori o validitate de 0,40 poate fi nesatisfăcătoare. Chiar dacă Ursula Şchiopu (1974) apreciază că o validitate de peste 0,25 face ca testul să dea predicţii superioare oricărei sondări empirice, asociaţiile profesionale ale psihologilor din America au dat în 1985, prin *Standards for Educational and Psychological Testing* o listă de probleme ce pot apărea în interpretarea coeficienţilor de validitate. Redăm câteva din ele:

- *Validitatea nu se măsoară, ci se deduce*. Deşi autorul testului trebuie să dea mai mulţi coeficienţi de validitate, indicând procedura urmată, utilizarea lui pe alte grupuri, sau în alte scopuri, necesită studii de validare suplimentare. Aceasta deoarece validitatea testului depinde simultan de scopul testării şi de populaţia căreia i se administrează. Toate aceste determinări ale validităţii care vin din studii ulterioare, desfăşurate în scopuri specifice şi pe alte eşantioane de populaţie, vor fi considerate doar dovezi ale unui anumit tip de validitate.

- Studierea validităţii testelor *angajează metode elaborate de analiză* (analiza de itemi, calculul corelaţiei, analiza factorială, regresia simplă şi multiplă) care combinate, dau *strategii de validare* mai bine definite. Utilizarea lor se face în funcţie de scopul şi de exigenţele beneficiarului.
- Strategiile de validare *se aplică odată cu construirea testului şi se repetă de câte ori este nevoie* pentru a obţine un test bun ca lungime, mod de aplicare, forţă discriminativă, uşurinţă în administrare, cotare şi interpretare, care alături de fidelitate şi validitate, dau o imagine mai adecvată despre valoarea sa.

Utilizarea testului impune procesul *repetării periodice* a cercetării validităţii, întrucât validarea unui test nu se încheie niciodată, datele nou acumulate din diverse domenii conducând la ameliorarea şi rafinarea sa, dar nu şi la o soluţie definitivă. Cu atât mai mult se pune problema reluării studiului validităţii unui test când i se schimbă formatul (prin aplicarea pe calculator, de exemplu), conţinutul, instructajul sau modul de cotare.

CAPITOLUL 8

ANALIZA DE ITEMI

8.1. Precizări terminologice

Într-o manieră similară cu cea din fizică, în care interesul iniţial pentru lumea macroscopică s-a recentrat pe cea microscopică, atomică şi subatomică, în psihodiagnostic această schimbare se manifestă prin mutarea centrului de greutate al preocupărilor de analiză de pe testul psihologic ca unitate de bază, pe analiza de itemi, adică pe cele mai mici unităţi care compun un test şi care pot fi analizate independent. Fundamentarea măsurării în psihologie se face prin psihometrie iar unitatea fundamentală a acesteia a devenit itemul. Evoluţia şi rafinarea domeniului testelor este astăzi tot mai dependentă de analiza de itemi, pentru că – pentru a-i cita pe Laveault şi Grégoire – „analiza de itemi seamănă cu repetiţia unei orchestre. Într-o orchestră instrumentele trebuie să cânte într-o manieră armonioasă.” [5]

O defiiniţie mai largă a itemilor este aceea că ei reprezintă „orice tip de enunţ din alcătuirea unui test obiectiv care urmează a fi cotat ca o valoare.”[6] Elementul fundamental al acestei definiţii este partea care stipulează că itemul va fi cotat ca o valoare cifrică, permiţând deci determinări cantitative ale unor caracteristici ale sale, dar acest lucru nu este specific doar testelor obiective, ci oricărui tip de test, inclusiv celor educaţionale. Merită subliniată şi precizarea lui Stan (2002): itemul este componenta informaţională a unui test „care se întâlneşte la chestionarul de personalitate, la testele de cunoştinţe şcolare, la testele verbale şi neverbale de inteligenţă. Nu trebuie deci confundate noţiunile de *sarcină* a testului cu cea de *item* al testului”, deosebirea fundamentală fiind aceea că „itemul are autonomie informaţională care permite efectuarea unor prelucrări statistice proprii asupra rezultatelor sale, ceea ce permite evidenţierea altor indicatori.”[7] Individualitatea sau autonomia funcţională a unui test este acea caracteristică a sa de a contribui la scorul total al testului şi de a produce rezultate care pot fi prelucrate independent sau intercorelate cu rezultatul final la test. De aceea cea mai bună definiţie a itemului este cea potrivit căreia acesta reprezintă *cea mai mică unitate informaţională a unui test care poate fi analizată independent.*

---

[5] Laveault, D., Grégoire, J. (2002). *Itroduction aux théoriex des tests en Psychologie et en scieces de l'éducation.* 2eme edition. Bruxelles: Éditions De Boeck Université, p. 199.
[6] Băiceanu, L. (2004). *Dicţionar ilustrat de psihologie englez – român*. Bucureşti: Editura tehnică, p. 191.
[7] Stan, A. (2002). *Testul psihologic. Evoluţie, construcţie, aplicaţii*. Iaşi: Editura Polirom, p. 259.

La modul foarte general, analiza de itemi a unui test sau chestionar are scopul de evalua fidelitatea şi validitatea fiecărei unităţi componente, pentru a surprinde în termeni cantitativi „cât de efectiv contribuie fiecare item la fidelitatea şi validitatea de ansamblu a testului" sau chestionarului respectiv.[8] În sens foarte strict, analiza de itemi nu se va confunda cu analiza itemului, deoarece scopurile lor sunt distincte. Analiza de itemi vizează evidenţierea dificultăţii itemilor, a puterii lor discriminative, a calităţilor psihometrice de tipul fidelităţii şi al validităţii, dar şi al altor caracteristici ca simetria şi boltirea (excesul), omogenitatea şi specificitatea sa ori raportul item-distractor. Analiza itemului se referă mai ales la procedurile corecte de prelucrare informaţională, la procedurile semantice de formulare sau de modificare a conţinutului itemilor, la limita de timp necesară rezolvării sale, la ambiguitatea formulării.

Analiza de itemi este greu de desfăşurat pentru testele de viteză sau pentru cele în care factorul rapiditate intervine masiv, cum ar fi testele de atenţie, testele de performanţă cu limită de timp, şi aceasta deoarece o condiţie a analizei de itemi este parcurgerea integrală a tuturor itemilor testului, de către toţi subiecţii, fără de care indicele de dificultate nu poate fi determinat. Deşi această condiţie nu este formulată explicit la testele de putere, în care factorul timp contează, problema se poate rezolva dându-se testul spre rezolvare în limita de timp prevăzută, când itemii se rezolvă în ordine, se marchează ultimul item pasat la expirarea timpului, după care testul se continuă până la rezolvarea sa în întregime de către toţi subiecţii.

Dificultatea majoră a anlizei de itemi este aceea că există o legătură evidentă între caracteristicile testului şi ale itemilor săi şi caracteristicile eşantioanelor care sunt măsurate cu ajutorul acestuia. Noua direcţie în psihometria modernă aspiră la degajarea parametrilor testului de caracteristicile eşantioanelor cercetate, dar aceasta complică destul de mult datele problemei. În cazul în care analiza se face plecând de la un singur eşantion, acesta ar trebui să aibă într-o cât mai mare măsură caracteristica de a fi reprezentativ, pentru a putea extrapola caracteristicile degajate de pe acest eşantion pe alte grupuri sau subgrupuri populaţionale pe care urmează a fi aplicat.

Analiza de itemi este o operaţie de neevitat în construcţia unui bun test sau chestionar, fie pentru a genera unităţi de măsură valide ale unui singur construct (unidimensionalitatea scalei de măsură), fie pentru a selecta itemii care au cele mai bune combinaţii dintre dificultate – forţă de discriminare şi omogenitate – specificitate. Aceasta este raţiunea pentru care analiza de itemi pleacă de la un set iniţial mai mare de itemi (cu până la o treime sau

[8] Reber, A.S. (1985). *Dictionary of Psychology*. London, New York etc.: Penguin Books, p. 378.

chiar dublu), din care vor rămânea cei care corespund cel mai bine exigenţelor de construcţie a unui test modern. Numărul minim de itemi care vor rămânea în urma acestei selecţii este de 10, sub care nici o scală nu mai discriminează suficient de nuanţat între diverse subcategorii, fiecare item având o pondere mult prea mare în scorul final. De regulă, scalele unidimensionale reţin între 20 şi 30 de itemi pentru a împăca criteriul economiei cu cel al sensibilităţii şi al forţei de discriminare a testului.

Elementul tehnic necesar pentru analiza caracteristicilor de dificultate şi discriminabilitate ale itemilor este *matricea lui Stern*[9] care plasează pe fiecare coloană itemii ce vor fi supuşi analizei şi pe fiecare linie câte un subiect, ultimele linii şi coloane fiind necesare sumarizării rezultatelor şi calculării unor indici. Analiza de pe coloană se numeşte *ipsatizare* (rezultatele tuturor subiecţilor la un anumit item), iar cea de pe linie *obiectivare* (rezultatele unui singur subiect la toţi itemii). Împărţirea scorului total pentru toţi subiecţii la numărul acestora dă o medie pe subiect, iar împărţirea scorului total pentru fiecare item la numărul de subiecţi dă o medie a reuşitelor. În concluzie, se poate determina o medie pentru fiecare subiect, o medie pentru fiecare item şi o medie pentru testul total, care este de fapt o sumă a mediilor tuturor itemilor.

8.2. Dificultatea itemilor

În cazul itemilor dihotomici – adică scoraţi cu 0 şi 1 – dificultatea unui item este dată de proporţia celor care au reuşit rezolvarea sa. *Indicele de dificultate* poate lua valori cuprinse între 0 şi 1, în prima situaţie itemul fiind foarte greu (majoritatea, sau chiar toţi subiecţii nu l-au putut rezolva), în a doua situaţie el fiind foarte uşor (majoritatea sau aproape toţi subiecţii l-au rezolvat). Dacă itemul a fost cotat pe o scală de mai multe puncte (polihotomică), indicele de dificultate se află după formula:

$$p = \frac{\sum x}{n}$$

în care Σx este suma scorurilor obţinute fiecare subiect iar n este scorul maxim. Pentru a avea o imagine mai clară asupra dificultăţii itemului, prin comparaţie cu un alt indicator care este media itemului, dăm tabelul de mai jos, în care sunt calculate mediile şi dificultăţile a 3 itemi scoraţi pe o scală de 5, 3 şi 1 punct.

[9] William Stern, 1912.

Aşa cum se observă din exemplul de mai jos, media itemului nu permite comparaţia directă a itemilor, deoarece numai pentru scala dihotomică cei doi indici - media şi procentul de reuşită - coincid. Pentru a fi direct comparabile, rezultatele trebuie aduse la acelaşi tip de scală, prin împărţirea mediei la valoarea maximă a fiecărei scale şi în felul acestea ele vor indica în acelaşi fel dificultatea itemilor, ca şi când toţi itemii ar fi scoraţi dihotomic.

Subiect	**Itemi**			
	Item (/5)	**Item (/3)**	**Item (/1)**	**Total (/8)**
1	4	3	1	8
2	5	2	0	7
3	1	0	0	1
4	3	1	1	5
5	4	2	1	7
6	2	1	1	4
7	2	3	0	5
8	0	3	1	4
9	5	2	1	8
10	1	1	1	3
11	3	1	0	4
12	3	2	1	6
Media (Σx/n)	33/12 = **2,75**	21/12 = **1,75**	8/12 = **0,67**	62/12 = **5,17**
Dificultatea p	2,75/5 = **0,55**	1,75/3 = **0,58**	0,67/1 = **0,67**	5,17/9 = **0,57**
Complement q (1-p)	1-0,55 = **0,45**	1-0,58 = **0,42**	1-0,67 = **0,33**	1-0,65 = **0,43**
Corecţie p'	**0,44**	**0,37**	**0,34**	-

Figura 8.1. *Mediile şi indicii de dificultate pentru 3 itemi scoraţi pe o scală de 5, 3 şi 1 punct.*

În exemplul de mai sus valoarea medie a dificultăţii celor 3 itemi este de 0,57, dar observăm că itemul cel mai dificil este pe scala de 5 puncte (p = 0,55), apoi cel pe scala de 3 puncte (p = 0,58), cel mai uşor fiind cel de pe scala dihotomică (p = 0,67). Dificultatea medie a itemilor sumarizaţi pe ultima coloană este de 0,57, valoare care poate fi considerată media itemilor întregului test, care se poate obţine fie împărţind media de 5,17 la 5+3+1 = 9, fie făcând o medie a dificultăţii celor 3 itemi (0,55+0,58+0,67)/3 = 0,60.

Aşa cum reiese din cele arătate anterior, itemul care se apropie ca proporţie p de rezolvare de valoarea 1 este unul extrem de uşor, deoarece majoritatea sau chiar toţi subiecţii îl rezolvă, pe când itemul al cărui p care se apropioe 0 este unul extrem de greu, fiind rezolvat de extrem de puţini subiecţi, poate de niciunul. Dificultatea medie a unui item este de 0,50, de regulă itemii aflaţi spre valori p extreme – 0 şi 1 – se recomandă să fie înlocuiţi, ei aducând puţine servicii psihodiagnostice testului, pentru a se reţine prioritar itemii de dificultate medie (p = 0,40 – 0,60). În scopul de a da o gradaţie mai fină scalei, limitele de acceptabilitate ale dificultăţii itemilor se plaseată de fapt între 0,20 şi 0,80.

Cunoaşterea dificultăţii itemilor este importantă din mai multe motive:

- permite eliminarea itemilor mult prea uşori ori mult prea dificili;
- permite selecţia şi organizarea itemilor pe scale ierarhice de dificultate, de la uşor la greu, acestea dând gradaţii cu atât mai fine cu cât numărul itemilor este mai mare şi diferenţele de dificultate dintre itemii progresivi sunt mai mici, p acoperind însă spectrul de la 0,20 la 0,80, sau chiar depăşind-o dacă se impune;
- permite crearea de seturi mari de itemi, cunoaşterea caracteristicilor privitoare la dificultate şi la indicele de discriminare, ca şi curba caracteristică a fiecărui item dând posibilitatea de a fi asamblaţi automat în examenul adaptat computerizat.

De notat că decizia de a elimina un item se poate lua mai uşor sau mai greu, deoarece acelaşi item care la o categorie superioară de vârstă este prea uşor devine greu sau foarte greu la o categorie mai mică de vârstă, itemul căpătând valenţe diagnostice noi funcţie de particularităţile eşantionului pe care se face analiza de itemi. Mai mult, un item foarte uşor, chiar dacă nu are calităţi psihometrice, dă încredere în sine subiectului şi îl ajută să intre în examen cu o reuşită, ridicându-i motivaţia pentru testare, pe când un item foarte greu poate distinge spre extrema excepţionalităţii. Aşadar păstrarea sau respingerea unor itemi se face în funcţie de utilitatea care se va da testului respectiv. Mai mult, problema dificultăţii itemilor este importantă în mod special la testele de randament sau la cele educaţionale (de cunoştinţe şi de achiziţii), pe când la testele de personalitate indicele de dificultate va fi înlocuit cu un *indice de proeminenţă*[10], capabil să evidenţieze măsura în care un item reliefează un anumit comportament rezultat din operaţionalizarea unei trăsături de personalitate.

Dintre factorii care pot afecta indicele de dificultate al unui test, Laveault şi Grégoire enumeră numărul de răspunsuri omise ca urmare a existenţei unei limite de timp şi probabilitatea reuşitei prin şansă/hazard. În prima situaţie, din cauza presiunilor legate de timp, un mare număr de persoane nu au putut rezolva rezolva un item amplasat de regulă spre zona finală a testului (atunci când itemii au fost asamblaţi ierarhic), indicele de dificultate ne mai reflectând acurat dificultatea itemului respectiv, pentru că dacă ar fi existat mai mult timp, cu siguranţă că el ar fi fost rezolvat de mai mulţi subiecţi. În acest caz indicele de dificultate măsoară de fapt două lucruri distincte, dificultatea itemului în cauză şi rapiditatea respondentului, caracteristici care nu întotdeauna se găsesc reunite la acelaşi subiect.

Atunci când dificultatea se calculează pentru itemi la care alegerea răspunsului se face dintre mai multe posibilităţi trebuie să se ţină cont de probabilitatea alegerii răspunsului bun

[10] Stan, op. cit., p. 2007.

prin şansă sau hazard, care este cu atât mai mare cu cât numărul variantelor de răspuns este mai mic. Astfel, la o variantă corectă din 8, răspunsul bun prin şansă are probabilitatea de 12 procente, la 1 din 6 de 17 procente, la 1 din 4 de 25 de procente şi la 1 din 2 de 50 de procente. În scopul de a corija indicele de dificultate, atunci când distractorii au şanse apropiate de a fi aleşi, Laveault şi Grégoirea propun următoarea formulă de corecţie:

$$p' = p - \left[\frac{1-p}{M-1}\right]$$

în care p este indicele de dificultate iniţial, p' este cel corijat, iar M este numărul variantelor de răspuns pentru itemul în cauză. Introducând în formula de mai sus datele pentru primul item din exemplul de la figura 8.1, p' = 0,55 - [(1 – 0,55)/(5 -1)] şi p' = 0,44. Pentru al doilea item, cotat pe o scală de 3 posibilităţi, p' devine 0,37 iar pentru itemul dihotomic corecţia ajunge la 0,34. Comparând indicile de dificultate iniţial cu cel corijat se observă uşor că corecţia este cu atât mai puternică cu cât numărul variantelor de răspuns este mai mic. Se degajă de asemenea concluzia că atunci când avem de comparat mai mulţi itemi, exprimaţi pe scale diferite ca număr de variante de răspuns, aducerea lor la formatul unic al unei scale dihotomice este o condiţie necesară, dar nu şi suficientă, căci pentru a avea o bază de comparaţie comună trebuie să operăm corecţia necesară prin formula de mai sus.

Câteva caracteristici importante ale itemului se determină plecând de la constatarea că există o legătură strânsă între dificultatea sa şi simetria distribuţiei sale. Astfel, itemii a căror dificultate este în jurul mediei (p = 0,50) sunt simetrici iar cei cu dificultăţi extreme sunt asimetrici, dar în mod diferenţiat: cei cu dificultate foarte mare (p < 0,10) au o asimetrie puternic pozitivă, pe când cei cu dificultatea foarte mică (p > 0,90) au o asimetrie puternic negativă. De aceea este mai gravă situaţia unui item uşor, eşuat, decât a unuia greu, eşuat, pe care oricum îl rezolvă puţini subiecţi. De aici concluzia că itemii uşori sunt discriminativi pentru subiecţii slabi sau pentru vârstele mici, pe când cei dificili sunt discriminativi pentru subiecţii foarte dotaţi şi/sau pentru vârstele mai mari, la testele care evidenţiază o ierarhie sau care au un indice genetic puternic (creşterea performanţei paralel cu vârsta până la un punct). De aici şi rolurile diferite alocate itemilor în funcţie de dificultatea lor, cei mai utili fiind cei de dificultate medie, care diferenţiază subiecţii între ei indiferent de scorul obţinut, fiind utili şi pentru discriminarea subiecţilor buni şi a celor slabi.

Un concept important, derivat din cele arătate mai sus, este acela de *varianţă a itemului*, care ne informează în legătură cu gradul de împrăştiere al performanţei la acel item în jurul unei medii, fiind în acelaşi timp şi un indicator mai grosier al puterii de discriminare a

itemului. Pentru itemii cu mai multe valori de răspuns (*polihotomici*), varianţa se determinaă prin metoda clasică, dar pentru o scală dihotomică există o formulă mult mai simplă pentru a o determina:

$$s^2 = pq$$

în care s^2 este varianţa itemului, p este indicele de dificultate iar q complementul său (1 - p). Pentru a avea o reprezentare mai clară asupra acestei probleme prezentăm mai jos o situaţie ipotetică, constând din 11 itemi propuşi spre rezolvare la 10 subiecţi.

	I T E M I											
Subiecţi	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**Total**
1	1	1	1	1	1	1	1	1	1	1	0	10
2	1	1	1	1	1	1	1	1	1	0	0	9
3	1	1	1	1	1	1	1	1	0	0	0	8
4	1	1	1	1	1	1	1	0	0	0	0	7
5	1	1	1	1	1	1	0	0	0	0	0	6
6	1	1	1	1	1	0	0	0	0	0	0	5
7	1	1	1	1	0	0	0	0	0	0	0	4
8	1	1	1	0	0	0	0	0	0	0	0	3
9	1	1	0	0	0	0	0	0	0	0	0	2
10	1	0	0	0	0	0	0	0	0	0	0	1
p	1,0	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,0	x=5,50
q	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0	σ=2,87
pq	**0,00**	**0,09**	**0,18**	**0,21**	**0,24**	**0,25**	**0,24**	**0,21**	**0,16**	**0,09**	**0,00**	Σx=55
Σpq	0,00 + 0,09 + 0,16 + 0,24 + 0,25 + 0,24 + 0,21 + 0,16 + 0,09 + 0,00 = 1,65											Σx^2=385

Figura 8.2. *Evoluţia varianţei itemilor în funcţie de indicele de dificultate p.*

Din tabelul de mai sus se poate observa faptul că cea mai mare varianţă o au itemii din zona centrală a indicelui de dificultate (la p = 0,5 s^2 = 0,25) şi, pe măsură ce ne depărtăm simetric spre zonele extreme, varianţa itemilor descreşte de asemenea simetric, de la 0,21 la 0,18, apoi la 0,09 şi 0. Aceata este încă un argument matematic puternic pentru a selecta în test itemii din zona centrală a dificultăţii, pentru că aceştia au forţa discriminativă cea mai mare, şi spre subiecţii slabi, şi spre cei buni. Itemii cu varianţă mică au funcţii speciale, fiind destinaţi fie discriminării subiecţilor slabi (p > 0,90), fie celor buni (p < 0,10), dar chiar şi în cadrul acestor categorii speciale forţa lor de discriminare este mai scăzută din cauza varianţei lor mai mici. Pe de altă parte, din figura de mai sus se observă că varianţa totală a testului ($2,87^2$ = 8,25) nu este egală cu suma varianţelor fiecărui item în parte (Σpq = 1,65).[11]

[11] În mod explicit, varianţa unui test este egală cu suma varianţelor fiecărui item, la care se adaugă dublul sumei valorilor fiecărei covarianţe posibile în ansamblul itemilor componenţi. Vezi Stan, op. cit., p. 216.

Pornind de la datele de mai sus pot fi calculaţi doi indicatori care regăsesc la nivel de item caracteristicile de ansamblu ale formei distribuţiei testului, şi anume simetria (*skewness*) şi excesul sau boltirea (*kurtosis*), după formulele:

simetria = $\alpha_3 = \frac{q-p}{\sqrt{pq}}$, în care α_3 este asimetria itemului;

boltirea = $\alpha_4 = \frac{1}{pq} - 6$, ecuaţie a cărei valoare normală este 3.

	I T E M I											
Subiecţi	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**Total**
p	1,0	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,0	n = 10
q	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0	x =5,50
pq	**0,00**	**0,09**	**0,18**	**0,21**	**0,24**	**0,25**	**0,24**	**0,21**	**0,16**	**0,09**	**0,00**	σ=2,87
α_3	**-**	**2,67**	**1,50**	**0,87**	**0,41**	**0**	**-0,41**	**-0,87**	**-1,50**	**-2,67**	**-**	Σx=55
α_4	**-**	**5,11**	**0,25**	**-1,23**	**-1,83**	**-2**	**-1,83**	**-1,23**	**0,25**	**5,11**		Σx^2=385

Figura 8.3. *Indicatori ai simetriei şi ai boltirii pentru tabelul precedent*.

Tabelul de mai sus evidenţiază foarte clar faptul că valorile lui p cuprinse între 0,70 şi 0,30 dau distribuţii relativ simetrice şi că asimetria este cu atât mai accentuată cu cât ne depărtăm mai mult spre extreme: pentru valori ale lui p > 0,80 asimetria este puternic pozitivă iar pentru p < 0,20 aceasta este puternic negativă. Acesta poate fi şi motivul pentru care Guthke, Bötcher şi Spring (1991, p. 230) fixează limitele inferioară şi superioară pentru selecţia itemilor unui test de randament la p = 0, 20, respectiv p = 0,80.

Răspunsul la întrebarea „cât de dificili trebuie să fie itemii unui test?" depinde, pe de o parte, de utilizarea testului, pe de altă parte de tipurile de răspunsuri sau de lungimea scalei. Aşa cum am mai arătat şi anterior, un test în care itemii se scorează adevărat-fals (sau 0 – 1) dă o probabilitate de 50% ca răspunsul să se dea prin şansă (aleator), iar când variantele de răspuns sunt 4, această probabilitate este de 25%. Deci probabilitatea p_i, care descrie dificultatea unui item, se plasează între 0 şi 1 (nici un subiect, respectiv toţi subiecţii îl rezolvă). În practică, dificultatea optimă a unui item se calculează în doi timpi:

1. Se calculează jumătatea diferenţei dintre 100% succes şi şansa de performanţă. Pentru un test cu 4 posibilităţi de răspuns, probabilitatea unui răspuns corect prin şansă (întâmplare) este de 25%, deci:

$$\frac{1{,}00 - 0{,}25}{2} = \frac{0{,}75}{2} = 0{,}375$$

2. Se adaugă la valoarea găsită probabilitatea răspunsului prin şansă:

0,375 + 0,25 = 0,625 (dificultatea optimă).

Iată câteva exemple:

- Test cu două variante de răspuns: $p_s = 0,50 + (1 - 0,50)/2 = 0,750$
- Test cu patru variante de răspuns: $p_s = 0,25 + (1 - 0,25)/2 = 0,625$
- Test cu şase variante de răspuna: $p_s = 0,167 + (1- 0,167)/2 = 0,584$
- Test cu opt variante de răspuns: $p_s = 0,125 + (1 – 0,125)/2 = 0,563$
- Test cu zece variante de răspuns: $p_s = 0,10 + (1 -0,10)/2 = 0,550$.

Alegerea dificultăţii optime a unui test trebuie să rezolve această problemă complicată: cu cât itemii au mai puţine variante de răspuns, cu atât mai mare este probabilitatea de a nimeri prin şansă răspunsul corect, ceea ce obligă la creşterea dificultăţii optime la un nivel foarte ridicat, de 0,75. Aceasta coboară progresiv cu mărirea numărului de variante de răspuns spre 0,50, dar cu cât acestea sunt mai multe, timpul de analiză pentru fiecare item creşte (sunt mai multe variante de luat în considerare), ceea ce obligă la reducerea numărului de itemi. Această reducere afectează simultan fidelitatea şi validitatea testului în ansamblul său, ajungând astfel la alternativa variante mai multe de răspuns versus mai mulţi itemi. Dilema se rezolvă alegând între 4 şi 8 variante de răspuns, aceasta fiind o zonă de echilibru între cele două extreme.

În rezumat, precizăm încă o dată că pentru majoritatea testelor nu este deloc indicat să avem itemi de dificultate egală (eventual cu dificultatea optimă), pentru că urmărim să avem itemi pentru o varietate a nivelurilor de dificultate, astfel încât testul să aibă o bună putere discriminativă în ansamblul său. Pentru cele mai multe dintre teste itemii au o dificultate ce acoperă spectrul de la 0,20 la 0,80, pentru a maximiza astfel diferenţele dintre subiecţi. În funcţie de *scopul* testării, pot fi concentraţi itemi preponderent dificili (sub 0,20), ca în cazul selecţiei unor candidaţi pentru o facultate, post sau meserie foarte pretenţioasă, sau itemi foarte uşori (selecţie indulgentă).

Pe de altă parte, cunoaşterea dificultăţii itemilor este fundamentală în *aranjarea lor în test într-o ordine ierarhică*. Itemii foarte uşori, de la începutul testului, ajută persoana să înţeleagă ce se cere de la ea şi, chiar dacă ei nu au valoare psihometrică probată, ajută la ridicarea moralului, inspirând încrederea în resursele proprii. Itemii foarte dificili sunt destinaţi extremei drepte a scalei, deci supradotaţilor sau celor de vârste mari, având utilitate pentru un număr mai mic de cazuri, ce apar doar pe eşantioane extinse ca volum.

8.3. Discriminabilitatea itemului

Deşi forţa de discriminare a unui item poate fi evaluată grosier şi după varianţa sa, într-un mod mult mai specific discriminarea itemului leagă destinul său de destinul de ansamblu al testului din care el face parte. Astfel, se consideră că un item este discriminativ atunci când există un bun paralelism între scorurile la test în ansamblul său şi scorul la itemul în cauză. Aceasta înseamnă că scorurile mari la test se asociază cu o rată de reuşită ridicată pentru itemul respectiv, după cum scorurile scăzute la test se asociază cu un nivel scăzut de reuşită la acest item.

Există mai multe metode de a studia dacă destinul itemului este solidar cu cel al testului în ansamblul său. *Indicele de discriminare*, pus în circulaţie de W.G. Findley în 1956, se mai cheamă şi metoda grupurilor de contrast sau a grupurilor extreme. Acesta este probabil şi motivul pentru care metoda s-a bucurat de o mare popularitate printre psihometricieni, dar este şi motivul pentru care ea este astăzi criticată, căci uşurinţa analizei se sprijină pe faptul că lasă pe dinafară aproape jumătate dintre subiecţi (46%). Metoda propune pentru comparaţia itemilor unui test două grupuri aflate la extremele seriei de variaţie a scorurilor, 27%[12] din cei care au dat răspunsurile cele mai bune la test fiind comparaţi cu ceilalţi 27% care au dat cele mai slabe răspunsuri, pentru a determina pentru fiecare item în parte dacă reuşita se asociază cu apartenenţa la grupul superior. Dacă vom nota cu p_s indicele de dificultate pentru grupul superior şi cu p_i pe cel pentru grupul inferior, formula indicelui de discriminare va fi:

$$D = p_s - p_i$$

Concret, indicele de discriminare se calculează astfel:

1. după scorurile la test se determină grupul superior, grupul inferior şi cel neutru;
2. se caută proporţia în care subiecţii din grupul superior şi din cel inferior au răspuns corect la fiecare item;
3. pentru fiecare item se scade proporţia răspunsurilor corecte ale grupului slab din proporţia răspunsurilor corecte ale grupului bun, ceea ce dă indicele de discriminabilitate, a cărui valoare poate varia între -1 şi +1.

Valoarea 0 semnifică egalitatea rezolvării de către cele două grupuri şi deci absenţa discriminării; valorile pozitive indică preponderenţa rezolvării itemului de către grupul superior, iar cele negative de către grupul inferior şi de aceea el trebuie eliminat fără discuţie,

[12] Proporţia respectivă a fost propusă de Kelley în 1939.

căci în acest caz destinul itemului nu este solidar cu al testului din care el face parte. Se apreciază că valorile lui D între:

0,00 - 0,10 indică absenţa discriminării;

0,11 - 0,19 valoare de limită a discriminării;

0,20 - 0,29 valoare scăzută a discriminării;

0,30 - 0,39 valoare bună a discriminării;

$0{,}40 \rightarrow$ item cu o foarte bună putere discriminativă.

Metoda presupune cel puţin 30 de subiecţi în grup, pentru a nu avea mai puţin de 8 subiecţi în grupurile de contrast.

Utilitatea şi popularitatea indicelui D provine şi din uşurinţa cu care el poate fi determinat manual, neluând în calcul decât ceva mai mult de jumătate dintre subiecţi. Mai mult, în pofida acestui fapt, el dă rezultate întru totul comparabile cu metode de lucru mai rafinate, cum ar fi cele corelaţionale. De aceea în opinia lui Laveault şi Grégoire el poate fi folosit mai ales pentru testele educaţionale, cu condiţia ca nunărul itemilor şi al subiecţilor să fie suficient de mare (peste 30) căci, în caz contrar, valoarea acestui item este artificial umflată, fiecare item acoperind o proporţie prea mare din varianţa totală. Autorii citaţi aduc şi un alt punct de vedere interesant: în testele educaţionale indicele D poate servi nu numai la selecţia celor mai discriminativi itemi, pentru a construi o scală puternică, ci şi ca metodă de analiză a distractorilor şi a erorilor. Pentru aceştia, indicele de discriminare trebuie să aibă o valoare negativă, dar el va fi informativ şi în legătură cu faptul dacă cheia răspunsurilor nu este eronată sau dacă răspunsul ales de grupul superior nu este realmente acceptabil. Dacă nu, se va putea cerceta motivul pentru care răspunsul a părut a fi acceptabil în mare măsură pentru grupul superior. După corectarea lucrărilor, chestionarea elevilor poate clarifica problema în cauză, iar rezolvarea ei duce la ameliorarea structurii itemului şi implicit a instrumentului.

8.4. Raportul dintre dificultatea şi discriminabilitatea itemului

Determinarea indicilor de dificultate şi de discriminabuilitate a itemului nu creează probleme tehnice insurmontabile, deoarece acestea au un algoritm de lucru foarte clar. Exceptând problema eşantionării, aceşti indici sunt dependenţi în mare măsură de populaţiile din care au fost obţinuţi. Problema cu adevărat importantă şi destul de complexă este selecţia itemilor celor mai buni dintr-un test ţinând cont corelativ şi simultan de ambele condiţii (dificultate şi discriminabilitate). Aceasta deoarece atât indicele D, cât şi valorile

corelaţionale ce vor fi prezentate ulterior, sunt influenţate de dificultatea itemului, care devine astfel condiţia bazală pentru selecţie, fiind de aceea pusă pe abscisă.

Situaţia paradoxală a constructorului unui test de randament sau de cunoştinţe este aceea că el trebuie să opereze în zona de acţiune a două condiţii care par reciproc exclusive: pe de o parte selecţia de itemi ce diferenţiază bine între subiecţi (şi aceasta presupune ca dificultatea lor să fie centrată pe 0,50), iar pe de altă parte el nu ar renunţa nici la itemii mai uşori (deoarece favorizează intrarea în probă şi discriminează în zona performanţelor slabe) şi nici la cei mai grei (discriminează supradotarea şi excelenţa). În aceste cazuri speciale, itemii prea uşori sau prea grei trebuie judecaţi cu atenţie din perspectiva rolului jucat de ei în raport cu scopul evaluării. „Analiza de itemi poate să ajute la punerea de întrebări pertinente în legătură cu rolul jucat de fiecare item, ca şi asupra mijloacelor celor mai bune de a ameliora calitatea unui instrument de măsură", apreciază Laveault şi Grégoire[13].

Metoda simplă a lui Findley a fost operaţionalizată de Ebel, care a oferit o modalitate practică de interpretare simultană a dificultăţii şi forţei de discriminare a itemilor, metodă ce a devenit extrem de populară datorită utilizării sale extensive şi îndelungate. Chiar dacă lasă aproape jumătate din cazuri în afara analizei, utilizarea ei actuală este încă foarte mare, mai ales în domeniul testelor educaţionale. În favoarea sa pledează nu numai uşurinţa aplicării, ci şi eleganţa soluţiilor diferenţiate pentru diversele combinaţii ale dificultăţii şi discriminabilităţii itemilor.

Pentru a exemplifica toate acestea, furnizăm mai jos un exemplu ipotetic care ne va ajuta să înţelegem modul în care operează selecţia celor mai buni itemi dintr-o scală.

Item	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	1	1	1	0	1	0	1	0	1	0
2	1	1	1	1	0	0	1	1	0	0
3	1	0	1	1	1	0	1	0	1	1
4	0	0	0	0	1	0	1	0	1	1
5	1	1	1	0	1	0	1	0	1	1
6	1	1	1	1	1	1	1	1	0	1
7	1	1	1	0	1	0	1	0	1	1
8	1	1	1	0	1	0	0	0	1	0
9	1	0	0	1	1	0	1	0	1	1
10	1	0	1	1	1	0	1	1	1	0
11	0	0	1	1	1	0	1	0	1	0
...										
29	1	1	1	1	0	1	0	0	0	0
30	0	0	0	1	1	1	1	1	1	1
31	0	0	0	0	0	1	1	1	0	0

[13] Idem, p. 216.

32	1	0	1	1	0	1	1	0	0	1
33	0	1	0	1	1	1	0	1	0	0
34	1	1	0	1	0	1	1	0	0	1
35	0	1	1	0	0	1	1	1	0	0
36	0	0	0	0	0	1	1	1	1	0
37	1	0	1	1	0	1	1	1	0	1
38	1	0	0	0	0	1	1	0	0	0
39	1	1	1	1	1	1	1	1	0	0
Σ	15	11	14	12	12	12	19	10	11	10
p	**0,8**	**0,1**	**0,2**	**0,4**	**0,5**	**0,5**	**0,9**	**0,3**	**0,4**	**0,6**
p_s	0,8	0,5	0,8	0,5	0,9	0,1	0,9	0,3	0,8	0,5
p_i	0,5	0,5	0,5	0,6	0,3	1,0	0,8	0,6	0,1	0,4
$\Delta=(p_s- p_i)$	**0,3**	**0**	**0,3**	**0,1**	**0,6**	**-0,9**	**0,1**	**-0,3**	**0,7**	**0,1**

Figura 8.4. *Analiza capacităţii de discriminare pentru 10 itemi ipotetici după metoda lui Findley.*

		p = d i f i c u l t a t e a i t e m i l o r											**Decizia**
		Grup puternic										**Grup slab**	
Δ = capacitatea discriminativă a itemilor	**1,0**												**Acceptarea itemilor**
	0,9												
	0,8												
	0,7					9							
	0,6						5						
	0,5												
	0,4												
	0,3				3					1			
	0,2												**Revizia itemilor**
	0,1					4		10			7		
	0	2	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**	**1,00**	**Respingerea itemilor**
	-0,1												
	-0,2												
	-0,3				8								
	-0,4												
	-0,5												
	-0,6												
	-0,7												
	-0,8												
	-0,9						6						
	-1,0												

Figura 8.5. *Interpretarea capacităţii de discriminare pentru cei 10 itemi anteriori după schema lui Ebel.*

Comentarii

- itemii 9 şi 5 trebuie reţinuţi, având caracteristici ideale: dificultate medie şi capacitate de discriminare foarte puternică;
- itemul 3 trebuie de asemenea reţinut: el este la limită ca dificultate, dar discriminează destul de bine;
- şi itemul 1 trebuie reţinut: deşi este foarte dificil, el se încadrează în zona p = 0,20 – 0,80 şi discriminează bine;

- itemii 4, 10 şi 7 sunt de revizuit: 4 şi 10 au dificultate optimă, dar din păcate discriminează slab, iar itemul 7 este foarte dificil şi discriminează slab;
- itemul 2 este de eliminat, odată pentru că este extrem de dificil şi în al doilea rând pentru că discriminează slab;
- itemul 8 va fi eliminat, pentru că deşi este de dificultate optimă, el discriminează negativ;
- la itemul 6 eliminarea este cel mai imperios necesară: deşi foarte bun ca dificultate, are o foarte mare capacitate de discriminare, dar negativă, fiind cel mai puternic în contratimp cum scorul general la test şi deci cel mai contraproductiv. Eventual el ar putea menţinut doar prin scorare inversă.

8.5. Indici corelaţionali ai discriminării itemilor

Ca şi indicele de discriminare D al lui Findley, metodele corelaţionale pot surprinde foarte bine paralelismul dintre destinul itemului şi cel al scorului total la test, în plus ele având şi avantajul de face să participe la analiză toate datele, şi nu doar o parte dintre ele. Mai mult, pentru fiecare coeficient de corelaţie se poate determina semnificaţia sa statistică, ceea ce dă mai multă precizie şi rigoare interpretării sale.

Dacă procedurile computerizate nu creează nici o problemă pentru determinarea lui r al lui Pearson, r_s (rho) al lui Spearman sau tau-b al lui Kendall, alte proceduri corelaţionale sunt mai laborioase şi mai pretenţioase. Acestea se referă la corelaţiile în care o variabilă sau ambele sunt măsurate pe scale non-numerice, discontinue (dihotomice sau polihotomice). În acest caz nu numai procedura de calcul este mai complicată, dar chiar selecţia celei mai potrivite metode trebuie făcută cu grijă. Laveault şi Grégoire (op. cit., p. 210) fac câteva recomandări sintetizate în figura de mai jos:

Scala de măsură	***Dihotomică***	***Dihotomizată***	***Continuă***
Dihotomică	Φ (Fi)	Φ_{bis}	r_{pbis}
Dihotomizată		$r_{tetrahoric}$	r_{bis}
Continuă			r, r_s (rho)

Figura 8.6. *Selecţia celui mai potrivit tip de corelaţie după tipul de scală.*

Menţiuni

O variabilă este *dihotomică* prin natura ei dacă nu are decât două stări posibile, cum ar fi masculin-feminin, ca apartenenţă de gen; ea este *dihotomizată* atunci când este adusă la doar două valori, deşi iniţial mai multe variante erau posibile, cum ar fi căsătorit-necăsătorit,

rural-urban, introvert-extravert, admis-respins etc., deşi pentru fiecare din acestea există şi alte variante intermediare: pentru starea civilă mai există şi variantele separat, divorţat sau văduv, pentru mediu şi varianta preurban, pentru extraversie/introversie şi varianta anbivert etc.; variabilele *continue* sunt cele de pe scalele de intervale sau de rapoarte, care pot lua în principiu orice valoare pe toată întinderea amplitudinii variabilei.

Pentru că metodele corelaţionale constituie un capitol distinct al statisticii, redăm condensat trei dintre coeficienţi de corelaţie prezenţi în diagrama de mai sus.

(1) $$r_{pbis} = \frac{\overline{x_p} - \bar{x}}{s_x}\sqrt{\frac{p_i}{q_i}}$$

unde $\bar{x}_p$ este media variabilei *x* (scorul total la test) a subiecţilor ce au reuşit la itemul *i*;

$\bar{x}$ este media variabilei *x* (scorul total la test) a tuturor subiecţilor;

s_x este abaterea standard a scorului la test;

p_i este proporţia subiecţilor care au răspuns corect la itemul *i*;

q_i este proporţia subiecţilor care au răspuns greşit la itemul *i*.

(2) $$r_{pbis} = \frac{\vec{x}_s - \vec{x}_i}{s_x}\sqrt{\frac{p}{q}}$$

unde $\vec{x}_s$ şi $\vec{x}_i$ reprezintă mediile grupului superior şi inferior;

S_x este abaterea standar a întregului grup;

p şi q reprezintă proporţia reusitelor la item şi complementul său

(3) $$r_{bis} = \frac{\vec{x}_s - \vec{x}_i}{s_x}\sqrt{\frac{p}{q}}$$

unde $\vec{x}_s$ este media grupului superior iar $\vec{x}_i$ media întregului grup;

s_x este abaterea standard a întregului grup;

y este este ordonata lui p transformată în note **z**; de fapt raportul $\sqrt{\frac{p}{q}}$ este furnizat de tabele speciale;

(4) $r_{bis} = \frac{\sqrt{pq}}{Y} \cdot r_{pbis}$, aceasta fiind ecuaţia de transformare între cei doi coeficienţi de corelaţie biserial şi punct-biserial.

Această mare diversitate de mijloace corelaţionale, adaugată la clasicul indice de discriminare al lui Findley ar putea deconcerta. Din fericire, există o foarte bună suprapunere şi concordanţă între rezultatele pe care ele le furnizează, exceptând valorile extreme ale dificultăţii itemilor, 0 şi 1, unde apar cele mai mari diferenţe şi discordanţe. Pentru aceasta furnizăm sfatul avizat al experţilor Crocker şi Algina[14]:

- toate metodele, corelaţionale sau noncorelaţionale, au o bună concordanţă în zona medie a indicilor de dificultate; deşi se poate utiliza foarte bine indicele D, o metodă corelaţională are avantajul suplimenar al testelor de semnificaţie care arată mai precis pragul de încredere al relaţiei descoperite;
- când indicii de dificultate se situează spre extremele 0 şi 1 trebuie folosită corelaţia biserială;
- acelaşi lucru este valabil şi atunci când se suspectează că eşantioanele viitoare pe care va fi aplicat testul diferă semnificativ de cel pe care acesta a fost aplicat iniţial;
- când se aşteaptă acelaşi nivel al aptitudunii la viitorul eşantion ca la cel iniţial este indicată corelaţia punct-biserială;
- când itemul şi variabila criteriu sunt dihotomizate, sau cotate în această manieră, trebuie utilizat coeficientul de corelaţie tetrahoric, cu atât mai mult cu cât itemul sau criteriul iau valori extreme.

8.6. Curba caracteristică a itemului (CCI)

Dacă într-un spaţiu cartezian pe abscisă se pun scorurile progresiv crescătoare la test iar pe ordonată probabilitatea la care examinaţii rezolvă fiecare item, obţinem o reprezentare grafică ce dă o imagine intuitivă uşor de interpretat în legătură cu calităţile discriminatoare ale itemului respectiv. Ea materializează mai puţin sintetic, dar mai util, raportul dintre performanţa la test şi performanţa la item pentru că indică şi zona pe care relaţia este bună, staţionară sau rea. Aşa cum o defineşte Gregory, CCI „este de fapt o idealizare matematică a relaţiei dintre probabilitatea unui răspuns corect şi măsura în care trăsătura măsurată este deţinută de respondent.” [15]

Există o diversitate de modele ale curbei caracteristice a itemului care se sprijină pe diferite funcţii matematice, cum ar fi teoria danezului George Rasch (1966), considerat a fi cel mai simplu şi elegant model.

[14] Apud Laveault şi Grégoire, op. cit., p. 215.

[15] Gregory, op. cit., p. 142.

Acesta se sprijină pe două presupuneri:

- itemii testului sunt unidimensionali, ei măsurând o singură trăsătură comună;
- itemii testului variază pe un continuum al nivelului de dificultate.

Atunci când abilitatea de a rezolva un item este distribuită gaussian, CCI are binecunoscuta formă de S a ogivei lui Galton. Deşi aceasta este preferată de majoritatea psihometricienilor, pentru că permite deducţii matematice despre trăsătura în cauză, selecţia itemilor după curbele lor caracteristice depinde în mare măsură de scopul testării. De exemplu, atunci când se doreşte construirea unui test de selecţie care să evidenţieze clar un punct de tăietură a scorurilor (*cutoff*) sunt de preferat curbele cu prag abrupt (vezi ilustrările de mai jos). Alteori scopul este urmărirea validităţii diferenţiale, caz în care itemii ce dau curbe net diferite după gen, etnie sau cultură sunt cei care vor fi eliminaţi sau revizuiţi, pentru a avea teste nesexiste sau corecte cultural.

Cel care şi-a pus problema utilităţii teoriei răspunsului la itemi, care stă conceptual în spatele curbei caracteristice a itemilor, a fost Nunnally[16] (1978). El a fost cel care s-a mirat de faptul că această teorie nu a fost adoptată mai demult în serviciul construcţiei testelor psihologice, dar tot el este acela care a dat răspunsul la această întrebare, atunci când a descoperit că presupunerea unidimensionalităţii nu se verifică pentru cele mai multe dintre testele pe care le-a luat în discuţie.

Conform lui Gregory, marea utilitate a curbei caracteristice a itemilor vine dinspre viitor, adică dinspre testarea adaptată computerizată, în care fiecare subiect răspunde la un set individualizat şi unic de itemi, scorat pe o scală uniformă şi unidimensională. Fără CCI, acest tip de examen adaptat computerizat nici nu ar putea fi imaginat. Dată fiind legătura dintre teoria răspunsului la itemi şi curba caracteristică a itemilor, rezervăm acestei probleme un capitol special.

Mai jos oferim imaginea grafică a curbelor mai multor tipuri de itemi, cu menţiunea că pe abscisă sunt trecute trei niveluri ale dificultăţii – uşor, mediu şi greu –, iar pe verticală de asemenea trei niveluri ale probabilităţii răspunsului la item – mic, mijlociu şi mare.

[16] Nunnally, J.C. (1978). *Psychometric Theory*. (2_{nd} ed.). New York: McGaw Hill.

CAPITOLUL 9

NORMELE ŞI ETALONAREA TESTELOR. CLASIFICAREA TESTELOR

9. Normele şi etalonarea testelor

9.1. Normele testelor

„Nici un test nu are norme pentru întreaga specia umană!" afirmă Anne Anastasi (1976, p. 90). Aceasta înseamnă că rezultatele la orice test au limite de valabilitate, generate de populaţia pe care el a fost aplicat şi în raport cu care s-a obţinut sistemul de referinţă reprezentat de normele sale (baremele sau etaloanele).

În psihodiagnoză noţiunea de *eşantion* are o triplă accepţiune:

- se referă la elementele de comportament măsurate de un test (definit de Anastasi ca „măsură a unui eşantion de comportament");
- are în vedere porţiuni din testul deja construit (un *item* poate fi considerat cea mai mică unitate, eşantion din test);
- are un înţeles statistic, pentru că se referă la populaţia pe care s-a aplicat testul în forma lui finală, în raport cu care se construieşte etalonul (eşantion normativ).

Aceasta din urmă este accepţiunea pe care o vom da conceptului de eşantion, necesar din două motive: aprecierea performanţelor unui subiect nu se poate face decât prin comparaţie cu un grup de referinţă, iar cum această comparaţie nu se poate face cu întreaga populaţie, din ea se extrage un eşantion reprezentativ, pe care se aplică şi se etalonează testul. De aici rezultă că „normele la un test nu au un înţeles absolut, universal sau permanent" (Anastasi, op. cit.) şi că pentru a fi relevante ele trebuie raportate la eşantionul de referinţă, „care conferă valabilitate social-istorică testului sau bateriei de teste" (Şchiopu, 1976, p. 151). Aceasta este şi raţiunea pentru care un test trebuie revizuit periodic.

Cea mai importantă caracteristică a eşantionului este *reprezentativitatea* sa, care înseamnă că el va fi construit ţinând cont de toţi factorii de variabilitate importanţi: vârstă, sex, mediu (rural–urban), la care se adaugă nivelul de instrucţie, pregătirea profesională, statutul socio-economic şi cultural, compoziţia etnică, distribuţia geografică şi orice altă caracteristică, atribut pe care populaţia de ansamblu o are şi care ar putea să genereze diferenţe în scorurile le test. Eşantionarea este un capitol important al Statisticii, care şi-a pus la punct procedee bine definite de alegere a sa pentru a-i creşte reprezentativitatea: metoda

cotelor, procedeul loteriei, al pasului, al eşantionării stratificate, multifazice, al eşantionului fix, metoda panel etc., tratate şi cu alte ocazii.

Deoarece tabelul de norme este relativ la populaţia avută în vedere, „nu se pot utiliza etaloane întocmite în alte ţări şi în alte condiţii, tabelul de norme are o valoare locală, cel mult naţională" (Radu, 1993, p. 381). *Mărimea* (volumul) eşantionului este foarte importantă: dacă este prea mare el generează o cantitate de muncă greu de acoperit, dacă este prea mic poate să lase pe dinafară populaţii sau caracteristici importante ale acesteia, ceea ce îi diminuează reprezentativitatea.

Raportul omogenitate-eterogenitate este o altă însuşite imporatntă a eşantionului, căci în funcţie de numărul şi de ponderea atributelor sale, pot fi construite subeşantioane populaţionale caracteristice. Acolo unde pot fi puse în evidenţă diferenţe legate de vârstă, sex, mediu sau nivel de instruire, se vor construi etaloane pe categoriile amintite, în cadrul fiecăreia celelalte elemente importante ce caracterizează populaţia trebuind să se regăsească în proporţii egale sau apropiate.

În cazul în care se intenţionează ca testul să aibă aplicabilitate foarte largă, naţională, atunci eşantionul normativ va fi suficient de mare şi de cuprinzător pentru a da valori stabile şi reprezentative. Pentru aceasta, populaţia va trebui definită în avans în termenii obiectivelor testului, după care, prin proceduri statistice specifice, va fi asamblat eşantionul pe care el va fi aplicat. În ţările în care există populaţii de mai multe tipuri (eterogenitate etnică), diferenţe foarte mari în standardele de viaţă sau de cultură, este de preferat să avem în vedere etaloane regionale sau locale, etaloanele naţionale fiind indicate aici doar pentru studiul opiniilor, atitudinilor sau intenţiilor de vot. Studierea *erorii de eşantionare* (care trebuie să fie cât mai mică) şi a distribuţiei scorurilor obţinute (care trebuie să fie unimodală) sunt repere statistice importante în construirea eşantionului.

Volumul eşantionului presupune exigenţa ca din fiecare clasă să se extragă cel puţin 300 de persoane (Kline, 1993), ceea ce înseamnă că populaţiile largi şi eterogene, generând multe variabile ce vor forma clase diferite, impun eşantioane atât de mari încât este preferabil să utilizăm etaloane pentru populaţii mai mici şi mai omogene, pentru care obţinem *norme locale*. Atunci când se pun în evidenţă diferenţe însemnate ale mediilor diferitelor subgrupuri chiar în cadrul acestora, ele pot fi utilizate în construirea unor etaloane de fineţe, căci permit raportarea mai nuanţată a fiecărui caz nou la sisteme de referinţă diferenţiate. Aceste subgrupuri pot fi formate pe vârste, sex, clasă şcolară, tip de curriculum, regiune geografică, mediu rural sau urban, nivel economic etc.

Normele locale, dezvoltate cel mai adesea de înşişi utilizatorii de teste, sunt avantajoase pentru că definesc mult mai strict subgrupele menţionate anterior. Mai mult, pe măsură ce testul este utilizat, el furnizează continuu date pentru subgrupuri tot mai bine definite, de unde şi posibilitatea de a construi etaloane tot mai diferenţiate, încât la limită se poate face spre exemplu predicţia pentru un elev în raport cu propriile lui rezultate anterioare. Măsura progresului unui subiect în timp, predicţia succesului în achiziţia şcolară sau a performanţelor într-un serviciu/muncă se bazează mult mai mult pe normele locale, decât pe cele larg reprezentative naţional. Acestea sunt extrem de necesare pentru că oferă cel mai larg şi mai stabil sistem de referinţă, necesar pentru a valida, de exemplu, examenele naţionale (capacitatea sau bacalaureatul, în ţara noastră). Deoarece costurile de obţinere a unor astfel de etaloane sunt foarte ridicate şi deoarece practicianul are de rezolvat probleme de mai mică anvergură şi mult mai specifice, etalonul local îi este mai accesibil şi mai util pentru că tezaurizează în timp şi valorifică propria sa muncă.

9.2. Etalonarea testelor

Aplicarea testului dă scoruri sub forma cotelor brute, care se analizează şi se tratează statistic pentru a obţine cotele standard, sau normele fixate în etalon. În acest scop se folosesc două tipuri de transformări (Albu, 1998, p. 246):

- transformări liniare, prin care se obţin *cotele standard* ***z*** şi *cotele standardizate* (păstrând nemodificată ordinea indivizilor în eşantion, adică repartiţia nemodificată a cotelor);
- transformări neliniare (de arie), prin care se modifică repartiţia cotelor brute, pentru a semăna cu o repartiţie uniformă, ca în cazul *quartilelor, decilelor* sau *centilelor,* sau cu una normală, ca în cazul *scalelor normalizate* cu 5, 7, 9 sau 11 trepte.

Dacă în primul caz se modifică unitatea divizionară a scalei pentru a se lua arii egale în tranşe succesive (câte 10% în cazul decilelor, 25% în cazul quartilelor şi 1% în al centilelor), în al doilea caz unitatea de scală este egală şi cea care diferă este aria porţiunii de sub curbă, corespunzătoare unei unităţi, aceasta în cadrul unei repartiţii care a fost normalizată (combină transformările liniare cu cele de arie).

Pe lângă aceasta, unele etaloane folosesc *vârste echivalente, vârste mintale* sau *clase echivalente* pentru variabile care au o creştere continuă în raport cu vârsta (înălţimea, greutatea, inteligenţa, vocabularul sau cunoştinţele la o materie şcolară).

Clasele echivalente, folosite mai ales pentru ciclul primar, în construirea testelor de cunoştinţe standardizate, exprimă pentru fiecare clasă şi lună de şcoală, mediana cotelor brute

la test, ceea ce permite reperarea unui scor particular al unui elev: citire 4,5 (nivel de clasa a IV-a, luna a cincea a anului şcolar), aritmetică 3,6, istorie 4 etc.

Vârstele echivalente dau fie vârste mintale, ca la testele de inteligenţă de tip Binet (unde raportul vârstă mintală/ vârstă cronologică, multiplicat cu 100, dă coeficientul de inteligenţă), fie vârste ale citirii sau socotitului, pentru testele de cunoştinţe. Se pot calcula astfel coeficienţi de inteligenţă (QI), de dezvoltare (QD, ca în bateriile Gesell) sau de educaţie (QE).

Cotele standard **z** arată cu câte unităţi reprezentate dintr-o abatere standard se distanţează o cotă brută a unui subiect la un test, comparativ cu media populaţiei de referinţă. Formula este:

$$z = \frac{x - m}{s}$$

unde rezultatul este cota **z** corespunzătoare scorului la testul *x*, iar *m* şi *s* sunt media, respectiv abaterea standard a populaţiei pe care s-a făcut etalonarea. Cotele standard *z* au avantajul unităţii de scară egale, exprimabilă în abateri standard (*s*), dar dezavantajul de a avea valori pozitive şi negative (cuprinse în intervalul –3 ,+3) şi de a cuprinde populaţii prea mari pe porţiunea centrală (68% din cazuri sunt cuprinse în intervalul de plus şi minus o abatere standard).

Rezolvarea practică o constituie transformarea liniară asupra cotelor *z* şi obţinerea *cotelor standardizate* după formula :

$$z' = M + sz$$

unde z' este cota transformată a cotei standard *z*, iar *M* şi *s* sunt media şi abaterea standard pentru cotele standardizate obţinute prin transformare. Valorile z' se rotunjesc la numere întregi şi proprietăţile lor sunt aceleaşi cu ale cotelor standard *z* , exceptând fireşte media şi abaterea standard, modificate prin transformare. Iată câteva tipuri de transformări (după Albu, p. 332):

	SCALA							
	N	**C**	**WP**	**L**	**T**	**Z**	**QI**	**H**
Media	3	5	10	10	50	100	100	50
Ab. sta.	1	2	3	5	10	10	15	14

Figura 9.1. *Tipuri se scale utilizate în psihodiagnoză*

Foarte utilizate în psihologia aplicată sunt quartilele, decilele şi centilele. Dacă primele dau o împărţire grosieră, în numai patru clase, centilele dau o ierarhie de fineţe a

rangurilor, din procent în procent, dar aceasta înseamnă teste lungi, cu mulţi itemi de putere discriminativă variabilă (între 0,30 şi 0,70 cel mai adesea), aplicate pe populaţii foarte mari (peste 100 de subiecţi). De aceea mai practice sunt decilele: „faţă de gradul de exactitate al măsurii psihologice, împărţirea în decile – ceea ce înseamnă individualizarea a 10 trepte de reuşită la o probă – reprezintă adeseori limita maximă", afirmă Radu *et al.* (1993, p. 376). De aceea vom ilustra operaţia de construire a unui etalon în decile, adică decilarea, prin care efectivul total se împarte în 10 clase, fiecare cuprinzând 10% din numărul total de subiecţi. Prin analogie, centilarea înseamnă 100 de clase, a 1% din efectiv fiecare.

Primul decil este acel reper sau valoare a variabilei care are înaintea sa 10% din ansamblul datelor ordonate, al doilea decil este elementul ce are înainte 20% din aceste date etc. Decilele sunt deci indicatori ai locului sau ai poziţiei pe scara considerată şi pentru a determina cele 9 repere ce vor delimita cele 10 decile, datele trebuie ordonate ierarhic. Pe prima linie vom avea valorile (cotele) brute ale variabilei test (x), pe a doua frecvenţele absolute (f_a) şi pe a treia frecvenţele relative cumulate (f_c). Iată rezultatele la testul *Raven Standard* (*PM 38*) aplicate pe 448 de concurenţi la un examen de admitere, utilizate pentru exerciţiul nostru de etalonare:

X	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**
f_a	0	0	1	0	1	3	0	0	1	0	1	0	0	1	0	3	1	2	3	4	7	3	2
f_c	0	0	1	1	2	5	5	5	6	6	7	7	7	8	8	11	12	14	17	21	28	31	33

X	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**	**57**	**58**	**59**	**60**
f_a	11	12	15	18	24	20	28	28	36	37	32	29	35	23	28	22	13	4	5	0
f_c	44	56	71	89	113	133	161	189	225	262	294	323	358	381	404	426	439	443	448	448

N = 448; $\bar{x}$ = 48,63; S = 6,12; mediana = 49,50.

Figura 9.2. *Distribuţia unor rezultate la Matricile avansate Raven*

a. Etalonul în quartile

Quartilul 1 (inferior) are înaintea sa 25% din subiecţi, adică (448·25)/100 = 112. Căutăm pe linia frecvenţelor cumulate (f_c) cea mai apropiată valoare de 112, care este 113, corespunzând cotei brute de 45. Quartilul 2 (mediana), împarte pe din două efectivul, deci are înaintea sa 50% din efectiv, adică (448·50)/100 sau 448/2 = 224. Valoarea cea mai apropiată de aceasta este 225, ceea ce corespunde cotei brute de 49. Quartilul 3 (superior) are înaintea sa 75% din efectiv, adică (448·75)/100 = 336. Valoarea cea mai apropiată este 353, corespunzând cotei brute de 53.

Etalonul va arăta astfel:

Quartile	Clase	Procente	
		Teoretice	Cumulate
Q_1	0 – 45	25	25
Q_2	46 – 49	25	50
Q_3	50 – 53	25	75
Q_4	54 – 60	25	100

Figura 9.3. *Etalon în 4 clase (quartile).*

b. Etalonul în decile şi centile (percentile)

Primul decil are 10% din efectiv înaintea sa, deci (448·10)/100 sau 448/10 = 44,80. Rotunjind, obţinem 45. Pe linia frecvenţelor cumulate valoarea cea mai apropiată de aceasta este 44, corespunzătoare cotei brute de 41, care devine reper pentru primul decil. Al doilea decil are 20% din efectiv înaintea sa, deci (448·20)/100 = 89,60, rotunjit 90, care pe linia frecvenţelor cumulate are valoarea cea mai apropiată 89, corespunzând cotei brute de 44, care devine reperul celui de al doilea decil. În mod corespunzător, reperele pentru celelalte decile sunt 46, 48, 49, 50, 52, 53 şi 55.

Etalonul în decile arată finalmente astfel:

Decile	Clase	Procente	
		Teoretice	Cumulate
10	0 – 41	10	10
20	42 – 44	10	20
30	45 – 46	10	30
40	47 – 48	10	40
50	49	10	50
60	50	10	60
70	51 – 52	10	70
80	53	10	80
90	54 – 55	10	90
100	56 – 60	10	100

Figura 9.4. *Etalonul în 10 clase (decile).*

Pentru că etaloane în centile întâlnim rareori, cel mai adesea se calculează reperele pentru punctele percentile 3, 5, 25, 75, 95 şi 97, celelalte valori (10, 20, 30 etc.) fiind respectiv primul, al doilea, al treilea decil etc. Modalitatea de calcul este similară: percentilul 3 are 3 procente din efectiv înaintea sa, deci (448·3)/100 = 13,44 rotunjit la 13, care pe linia frecvenţelor cumulate are cea mai apropiată valoare 14, cu cota brută corespunzătoare 35 (reperul pentru percentilul 3). Similar, percentilul 97 este (448·97)/100 = 434,56, rotunjit 435, care are ca reper scorul brut 57.

c. Etaloanele în clase normalizate

Acest tip de etalonare se realizează în clase care au efective conforme distribuţiei normale, ceea ce presupune că distribuţia originală va fi normalizată prin folosirea de clase ce au proporţii derivate din proporţiile unei curbe gaussiene normale. Se folosesc mai multe astfel de etaloane, care pot avea 5, 7, 9, 10 sau 11 clase, cu următoarele distribuţii procentuale:

Etalon în 5 clase: 6,7; 24,2; 38,2; 24,2; 6,7 (%)

Etalon în 7 clase: 4,8; 11,1; 21,2; 25,8; 21,2; 11,1; 4,8 (%)

Stanine = 9 clase: 4,0; 6,6; 12,1; 17,5; 19,6; 17,5; 12,1; 6,6; 4,0 (%)

Stens = 10 clase:

Note C = 11 clase: 3,6; 4,5; 7,7; 11,6; 14,6; 16,0; 14,6; 11,6; 7,7; 4,5; 3,6 (%).

Pentru că etalonarea în stanine a cunoscut o largă extindere după cel de al doilea război mondial (ea dă 9 clase ce pot fi grupate între ele câte 3, fiecare cu câte 3 gradaţii; cu excepţia staninei 1 şi 9, au unităţi de scară echivalente, deci sunt aditive), vom exemplifica construirea unui etalon plecând tot de la datele anterioare. Stanina 1 are sub ea 4% din efectiv, deci (448·4)/100 = 17,92, rotunjit 18, care are pe linia frecvenţelor cumulate valoarea cea mai apropiată 17, căreia îi corespunde cota brută de 36. Similar stanina 2 are înaintea sa 4,0+6,6 = 10,0 procente din efectiv, adică (448·10,6)/100 = 47,468 rotunjit la 47, care pe linia frecvenţelor cumulate are cea mai apropiată valoare 41.

Etaloane de calitate se obţin atunci când volumul eşantionului depăşeşte 100 de subiecţi, selecţionaţi după criterii (de regulă nu mai mult de 4) de vârstă, sex, mediu, profesie etc., pentru a fi în acelaşi timp omogene în raport cu criteriul şi eterogene în aceeaşi proporţie cu populaţia mai mare din care a fost extras, adică pentru a fi reprezentativ. Iată etalonul:

Stanine	Clase	Procente:	
		Teoretice	Cumulate
1	0 – 36	4,0	4,0
2	37 – 41	6,6	10,6
3	42 – 45	12,1	22,7
4	46 – 48	17,5	40,2
5	49 – 50	19,6	59,8
6	51 – 53	17,5	77,3
7	54 – 55	12,1	89,4
8	56	6,6	96,0
9	57 - 60	4,0	100,0

Figura 9.5. *Etalonul în 9 clase (stanine).*

Trebuie precizat că de foarte mare utilita sunt două tipuri de scale standardizate, scorurile T (introduse de MCCall în 1922) şi scorurile H, introduse de Hull, ale căror formule de transformare a scorurilor brute în scoruri standard sunt următoarele:

$$T = 50 + \frac{10}{s}(x - M)$$

respectiv

$$H = 50 + \frac{14}{s}(x - M).$$

9.3. Clasificarea testelor

În binecunoscuta sa lucrare *Psychological Testing*, Anne Anastasi (1976) tratează testele psihologice după trei mari subdiviziuni:

1. ***teste de nivel intelectual general***, în care include testele individuale, testele pentru populaţii speciale şi testele de grup;
2. ***testele pentru abilităţi separate***, cu trei subdiviziuni majore: teste educaţionale, ocupaţionale şi clinice;
3. ***testele de personalitate***, care includ inventarele autoaplicate, testele ce măsoară interese, atitudini, valori şi tehnicile proiective.

Aiken (1997) tratează ***evaluarea abilităţilor*** (inteligenţa şi abilităţile speciale) şi a ***personalităţii*** (interese, atitudini, valori; observaţia, interviul şi ratingul; inventarele şi tehnicile proiective).

Kaplan şi Saccuzzo (1993) disting ***tehnicile interviului***, testele de ***abilităţi mintale***, alte teste de ***abilităţi individuale sau de grup***, apoi teste ***pentru alegerea carierei***, teste de ***personalitate structurate***, teste de ***personalitate proiective***, la care adaugă procedurile de ***evaluare a comportamentului***, de măsură a ***anxietăţii şi a stresului***, testarea ***stării de sănătate*** şi ***testele computerizate***.

Ion Holban (1973) propune o clasificare ce cuprindea testele de ***randament***, de ***dezvoltare mintală***, de ***aptitudini***, de ***cunoştinţe***, de ***comportament***, de ***alegere,*** teste ***proiective***, **sociometrice**, degajând şi alte criterii de clasificare.

Ursula Şchiopu (1976) propune ca mari categorii de tehnici psihodiagnostice testele ***de dezvoltare***, cele de ***inteligenţă***, de ***aptitudini***, de ***personalitate*** şi de ***examinare psihologică a sociabilităţii***.

Fiecare clasificare configurează în alt mod domeniul psihodiagnozei, care uneori se rezumă la testare (*testing*), alteori şi la evaluare (*assessment*), putând include şi toate metodele clasice utilizate în vederea diagnozei (anamneza, metoda biografică, convorbirea, observaţia, ancheta, interviul, chestionarele sau metodele sociometrice). Deşi observăm că multe dintre clasificările enunţate mai sus se suprapun în bună măsură, este necesar să delimităm câteva criterii clare de clasificare, care ne vor ajuta să circumscriem mai bine domeniul testării psihologice şi să înţelegem mai nuanţat diferenţa specifică dintre tipurile de teste. Deoarece este greu de ierarhizat ponderea fiecărui criteriu şi deoarece graniţele dintre ele sunt relative, ordinea lor nu este neapărat cea pe care o vom adopta în continuare.

9.3.1 Clasificarea testelor după modul de administrare: teste individuale şi de grup

- ***Testele individuale***, foarte apropiate de metoda clinică, sunt mari consumatoare de timp, dar ele permit stabilirea unui contact mai bun cu subiectul, ce va fi observat mai uşor, informaţia dată de test putând fi suplimentată semnificativ cu alte date care sporesc acurateţea diagnosticului.
- ***Testele de grup (colective)****:* în cazul lor examenul este mai strict standardizat (pentru a controla toate elementele ce ar putea denatura rezultatul final, în special neînţelegerea instructajului şi frauda), adună o mare cantitate de date într-un interval scurt (deci sunt foarte economice), dar sărăcesc examenul de date importante privind motivaţia, implicarea în probă, oboseala, anxietatea, boala sau plictiseala. Sunt foarte indicate în examenele de selecţie primară, de cunoaştere a opiniilor sau a atitudinilor. Pentru consiliere, ghidare vocaţională, orientare sau în clinică sunt de preferat abordările individuale, datorită plusului lor de calitate.

9.3.2. Clasificarea testelor după timpul de execuţie: teste de viteză şi teste de randament

- ***Testele de viteză*** pleacă de la presupunerea că există o corelaţie între viteza de lucru şi dezvoltarea unei caracteristici (atenţia, de exemplu). Sarcinile nu sunt organizate ierarhic, având acelaşi grad de dificultate, dar testul are un număr foarte mare de itemi, pentru a nu putea fi integral rezolvat de nici un subiect. Evident, primează aspectul cantitativ.
- ***Testele de randament*** au itemii de dificultăţi diferite, aranjaţi ierarhic, timpul fiind suficient de mare pentru a permite parcurgerea lor de către toţi subiecţii. Cota brută o

reprezintă numărul de probleme corect rezolvate (primează deci aspectul calitativ al rezolvării).

Cele mai multe teste combină aceste două criterii, cum este subtestul *Cuburilor WISC,* unde modelele de realizat sunt progresive ca dificultate (utilizează patternuri caroiate, apoi fără caroiaj, modele de 2x2, apoi de 3x3 cuburi), iar punctele rezultă din rezolvarea corectă a modelului şi din bonificaţiile de timp (de viteză, de fapt).

9.3.3. Clasificarea testelor după modul de executare a sarcinii de către subiect: în teste orale, scrise (tip creion-hârtie) şi de performanţă

- ***Testele orale*** cer răspunsuri verbale (informaţii, vocabular, completare de propoziţii, comprehensiune, asociere verbală etc.).
- ***Testele scrise*** cer un răspuns scris (care poate fi şi desenat), dat pe foaia de test sau de răspuns. Proliferarea lor extraordinară, dată de comoditatea soluţiei, tinde astăzi să fie blocată de extinderea examenului cu ajutorul calculatorului, care afişează pe ecran problema, soluţia fiind indicată prin tastă, mouse sau creion optic. Mai mult, calculatorul scanează foi de răspuns, scorează, analizează răspunsul şi furnizează o foaie de profil, având un rol esenţial şi în re-etalonarea progresivă a testului.
- ***Testele de performanţă*** presupun materiale, instrumente sau aparate, dispuse în condiţiile speciale de laborator, necesită o supraveghere atentă şi rareori permit examene colective.

În acest context, distincţia ***teste verbale–nonverbale***, deşi des utilizată, nu este operantă, pentru că este greu de specificat până la ce nivel intervine sau nu factorul verbal. Testele la care există preponderent răspunsuri motorii, construite cel mai adesea pentru a fi „*free culture*", adică independente de limbă, sunt destinate imigranţilor, handicapaţilor cultural, hipoacuzicilor ori minorităţilor etnice, adică cei cărora factorul verbal poate să le modifice negativ performanţa finală.

9.3.4. Clasificarea testelor după modul de procesare implicat: teste de eficienţă şi teste de personalitate

Fiecare categorie are mai multe subdiviziuni:

- ***Testele de eficienţă*** se referă la aptitudini generale (inteligenţă, memorie) şi speciale, la cunoştinţe, la probele de lucru şi la testele situaţionale.
- ***Testele de personalitate*** includ chestionarele autoaplicate, tehnicile proiective şi testele obiective de personalitate.

Dacă testele de eficienţă încearcă să surprindă performanţa maximă a unei persoane, cele de personalitate vor să surprindă performanţa ei tipică (Corsini, 1994), primele având itemi de genul adevărat–fals, prin care se obţine un scor numeric şi o ierarhizare a indivizilor, în timp ce la a doua categorie răspunsurile nu sunt dihotomice (bune/rele), deci nu permit o ierarhizare, ci o raportare tipologică.

Teste de eficienţă sunt la rândul lor:

- ***Testele de aptitudini*:** conformându-se unui model evolutiv şi ierarhic asupra inteligenţei, sunt *teste de aptitudine generală* (*inteligenţă*), care măsoară un potenţial general, teste de *aptitudini speciale* (muzică, mecanică, sport, arte etc.) şi teste de *aptitudini multiple,* care cuprind într-o singură baterie mai multe feluri de aptitudini, definite factorial, ce pot interveni în domenii specifice (dexteritate, inteligenţă spaţială, tehnică, manualitate etc.).
- ***Testele de cunoştinţe*** măsoară nu un potenţial (ca cele de inteligenţă), ci nivelul efectiv atins în achiziţia de informaţii sau de deprinderi specifice unui domeniu (matematică, istorie, informatică etc.). Pentru că cel mai bun predictor asupra a ceea ce poate învăţa cineva este ceea ce el deja a învăţat până la un moment dat, testele de cunoştinţe au în vedere în mod expres experienţa anterioară de învăţare. De fapt, cea mai bună predicţie este dată de folosirea concomitentă a testelor de aptitudini generale (inteligenţă) cu cele de cunoştinţe, pentru că combinarea lor este mai informativă (indică şi ceea ce este potenţial, dar şi ceea ce este dobândit prin învăţare). Unele dintre testele de cunoştinţe sunt de tip *informativ* (elaborate de profesorii înşişi pentru a determina nivelul unei clase), sau *standardizate* (elaborate de specialişti, destinate utilizării pe scară largă, având un manual care indică strict condiţiile de standardizare în aplicare, cotare, raportare la norme elaborate pe eşantioane reprezentative şi acoperind o mare varietate de conţinuturi, ce le fac foarte recomandabile pentru evaluări sumative).
- ***Probele de lucru*** sunt extrem de indicative pentru evaluarea nivelului atins într-o profesie, activitate, dependentă de formarea unor deprinderi motrice, intelectuale sau rezultate dintr-o experienţă. Unele dintre elementele implicate într-o muncă devin foarte diagnostice pentru rezultatele ce se vor obţine ulterior în munca respectivă. Înaintea unui program de instruire, deseori se face o astfel de evaluare care, când este bine concepută, are validitate predictivă şi face o mare economie de timp şi bani, programele putând dura luni sau ani.

- ***Testele situaţionle*** „sunt echivalentele probelor de lucru în cazul profesiilor administrative" (Albu, 1998, p. 74), deosebirea majoră fiind că în testul situaţional se regăseşte sarcina de lucru în totalitatea ei, şi nu numai anumite elemente. În afară de aceasta, ea nu cuprinde numai deprinderi (*skills*) ci şi variabile de personalitate şi sociale (emoţii, opinii, stil personal, aptitudinea de lider, etc.). Scopul lor este adesea deghizat, ele fiind construite ca simulări ale unor situaţii reale, pe care cel în cauză trebuie să le rezolve. Tehnica „in-basket" (*coşul de lucru*) simulează condiţiile de lucru ale unui manager, director de firmă, care trebuie să răspundă la toate solicitările cuprinse în „coş": telefoane, decizii, scrisori, directive, interviuri etc. Capacitatea de conducere şi stilul personal în lucrul cu oamenii se evidenţiază astfel printr-un test în care „itemii" sunt de aceeaşi natură şi complexitate ca munca efectivă. Evident că aceste teste se organizează după faza interviurilor preliminare şi după examenele psihologice cu teste clasice, prin care numărul candidaţilor cu şanse de a fi selectaţi pentru post se reduce drastic.

Teste de personalitate sunt:

- ***Chestionarele de personalitate*****.** Acestea pot fi unifazice sau multifazice, după cum măsoară una sau mai multe dimensiuni de personalitate (extraversie, anxietate, nevrotism, maturizare emoţională, forţa Eului, masculinitate-feminitate, tendinţe patologice etc.) După conţinut, ele pot fi *de adaptare,* caz în care se numesc *inventare de personalitate,* indicative pentru integrarea socială a propriei conduite sau pentru starea de sănătate. În afara acestora, chestionarele pot măsura *interesele* (foarte importante în orientarea profesională), *atitudinile, opiniile* (fundamentale în anchetele sociologice), sau *valorile*. Uşor de administrat şi de cotat, chestionarele de personalitate au o largă utilizare, căci se bazează pe supoziţia (discutabilă!) potrivit căreia „cea mai bună cale de a afla ceva despre un individ este de a-l întreba direct" (Hammond, 1995). Sunt multe dovezi că între sinele social şi cel psihologic, între sinele real şi cel ideal pot exista distanţe mari sau conflicte (surse de inadaptare socială) şi, mai mult, dacă situaţia de examinare are miză (angajare, selecţie profesională), răspunsurile pot fi trucate intenţionat (răspunsuri „de faţadă").
- ***Tehnicile proiective*** – nu sunt teste în sens strict, pentru că au un grad de standardizare şi etalonare redus şi, în consecinţă, calităţi psihometrice foarte discutabile, de unde şi rezerva multor psihometricieni faţă de acestea. Ele urmăresc sesizarea nu a ceea ce este comun tuturor oamenilor (planul nomotetic), ci aspectele idiografice, adică ceea ce este profund individual şi specific unei singure persoane, fiind calea cea mai importantă pentru

a ajunge la un diagnostic diferenţial. Cunoaşterea acestor factori unici, numită diagnoza la nivel intim, de adâncime, structural, plasează aceste probe în afara posibilităţilor de etalonare, de unde şi numele de *tehnici* şi nu de teste proiective. Au o utilizare foarte frecventă în clinică (în psihiatrie cu precădere), dar sunt foarte răspândite şi în domeniul judiciar, în şcoală sau transporturi, acolo unde se pune problema surprinderii unităţii, integrităţii sau sănătăţii mentale a persoanei. De multe ori ele premerg psihoterapia sau îi evaluează rezultatele. Sarcina subiectului are caracter foarte general şi vag precizat, el trebuind să spună ce-i trece prin minte, uneori o poveste despre o fotografie, imagine, desen, pată de cerneală, elementele inductoare fiind nişte stimuli nestructuraţi, vagi, ambigui. În construcţia relatării sale, subiectul va folosi materialul pe care îl are la dispoziţie, stocat în fondul aperceptiv al memoriei sau în inconştientul său, de unde şi numele de *probe proiective*. Pentru a le desemna se mai utilizează şi alţi termeni: teste aperceptive (ca TAT, CAT, Symonds), teste nestructurate sau de dinamism. Două sunt problemele de fond ale acestor tehnici: subiectivitatea accentuată a interpretării (validitate concurentă şi încrucişată reduse) şi timpul extrem de lung necesar formării unui psihodiagnostician bun, specializat în domeniu. Dacă la acestea adăugăm timpul mare, cerut de aplicare şi interpretarea rezultatelor, faptul că sunt folosite doar scale nominale (raportarea persoanei la o tipologie sau categorie) şi faptul că măsoară date ce au o mare fluctuaţie de moment (dependente de factori dispoziţionali), vom înţelege de ce entuziasmul iniţial pentru aceste tehnici s-a mai temperat, ele rămânând însă, pentru unele zone, singurele instrumente diagnostice dezirabile.

- ***Testele obiective de personalitate*** sunt, ca şi cele proiective, teste indirecte pentru că-şi deghizează scopul, dar sunt numite obiective pentru că folosesc stimuli structuraţi (activităţi simple, manipulabile experimental), din execuţia cărora se deduc aspecte relevante ale personalităţii, cu o mult mai mare concordanţă între evaluatori, comparativ cu cele proiective. Modificarea vigilităţii (*arous-ului* cortical) la introverţi şi extraverţi este extrem de diferită, primii având o descreştere accentuată, ceilalţi o creştere progresivă, paralel cu concentraţia de cofeină. Mulţi constructori de chestionare (inventare) de personalitate s-au lăsat seduşi de ideea găsirii unor validări externe a acestora prin teste obiective, printre aceştia numărându-se Eysenck sau Cattell.

9.3.5. Clasificarea testelor după constanţa conţinutului de la o administrare la alta: teste cu conţinut fix şi teste cu conţinut variabil

Dacă cele cu ***conţinut fix*** pot fi pot fi considerate teste în sensul clasic (toate persoanele testate parcurg aceiaşi itemi, prezentaţi în aceeaşi formă), cele cu ***conţinut variabil*** conturează un mare domeniu (uriaş) al testelor adaptate, administrate, scorate şi prelucrate computerizat. În acest caz nu mai există un test unic, ci o bancă de itemi, ale căror caracteristici au fost determinate din administrări anterioare (ansamblu larg de itemi calibraţi ca discriminabilitate şi dificultate, vezi Havârneanu, 2000).

Testele adaptative computerizate sunt construite pe baza teoriei răspunsului la itemi, selecţia itemului administrat fiind în funcţie de caracteristicile conjugate ale acestuia şi ale subiectului examinat, pentru a maximiza informaţia despre el, în condiţiile unei testări mai eficiente şi mai rapide. Algoritmul cuprinde o secvenţă care începe cu un item de dificultate medie care este administrat, se evaluează răspunsul în raport cu variabila test apoi, funcţie şi de răspunsurile anterioare ale subiectului, pentru itemii care vor urma se decide continuarea sau stoparea testării, dacă s-a atins un nivel de încredere specificat, sau timpul ori numărul de itemi de administrat a fost epuizat. Dacă la început testele adaptative computerizate măsurau doar aptitudini şi cunoştinţe (itemi omogeni relativ la construct), acum ele tind să se extindă spre măsurarea de variabile necognitive (opinii, atitudini, valori sau interese).

9.3.6. Clasificarea testelor după modul de cotare

Există teste ***obiective*** şi ***subiective***, distincţie care deja s-a degajat prin compararea tehnicilor proiective (cotare subiectivă) cu cele cognitive sau de cunoştinţe (obiective). Obiectivitatea în cotare este o importantă calitate psihometrică a unui test, care presupune ca sarcina să fie structurată şi răspunsurile bune să fie alese dintr-un număr fix de răspunsuri preformate. Ea se obţine şi când numărul itemilor este mare şi scorul la test are o formulă de calcul clar specificată.

9.3.7. Clasificarea testelor după modul de interpretare a scorurilor

- ***Teste normative***, unde scorul individual îşi dezvăluie semnificaţia prin raportarea la etalonul obţinut pe un eşantion reprezentativ statistic şi omogen.
- ***Teste centrate pe criteriu***, unde se stabileşte un anumit nivel acceptabil al performanţei, în funcţie de care subiectul este admis sau respins. Acestea sunt utilizate fie în selecţia primară (apt/inapt pentru conducerea auto), fie în şcoală, în ciclul primar, când criteriul

arată cât de mult s-a apropiat elevul de obiectivul educaţional scontat şi mai puţin de un anumit loc în ierarhia clasei.

- ***Teste idiografice***, în care se construieşte un instrument diagnostic doar pentru un singur individ, sau scorul său la un chestionar devine criteriu de evaluare a unui program psihoterapeutic, prin pre-test şi post-test.
- ***Teste ipsative,*** care cer subiectului să aloce o resursă personală limitată mai multor domenii, urmărind compararea între domenii sau între scorurile obţinute la mai multe din scalele testului.

9.3.8. Clasificarea testelor după numărul variabilelor implicate

- ***Teste analitice*.** Acestea s-au impus tot mai mult datorită dezvoltării metodelor de analiză factorială, care permit selecţia de itemi omogeni în raport cu un construct. Ele au la bază principiul din fizică al izolării fenomenului de investigat, ceea ce este relativ dificil la nivelul psihicului, căci el tinde, prin evoluţia sa, spre o unificare progresivă prin „largi sisteme de integrare" (Allport, 1937).
- ***Teste sintetice*.** Ele pun persoana în condiţiile de solicitare complexă, permiţând identificarea unor parametri aptitudinali diferiţi, concomitent cu surprinderea (prin observaţie), a unor trăsături sau chiar factori mai generali de personalitate.

9.3.9. Clasificarea testelor din punctul de vedere al procesului psihic investigat

Din acest punct de vedere şi în prelungirea unei perspective atomist-facultaţioniste în psihologie, tot mai puţin utilizată astăzi, există teste de ***atenţie***, de ***percepţie***, de ***memorie***, de ***inteligenţă***, de ***creativitate***, de ***personalitate*** etc. Este de ajuns să arătăm cum inteligenţa pune în mişcare toate rezervele psihice în scopul adaptării şi cum personalitatea tinde spre integrări succesive, spre a evidenţia relativitatea acestui criteriu, care păstrează o utilitate pur didactică.

Notă: Subcapitolul „Clasificarea testelor" a fost elaborat în principal pe baza capitolului 4 (*Clasificări ale testelor psihologice*) din lucrarea Monicăi Albu "*Construirea şi utilizarea testelor psihologice*", Cluj-Napoca, 2000, pp. 65 – 92, pe care îl recomandăm pentru detalieri şi o înţelegere mai analitică.

CAPITOLUL 10

DEZVOLTAREA SCALELOR PSIHOLOGICE

10.1. Introducere

O bună parte din activitatea practică a psihologului se sprijină pe diagnoza unor aspecte cantitative ale aptitudinilor şi dimensiunilor definitorii pentru personalitate. În acest scop el utilizează instrumente ale căror calităţi psihometrice trebuie construite şi evaluate în cunoştinţă de cauză. În plus, pentru unele dimensiuni psihologice, pedagogice sau sociologice specialistul trebuie să dezvolte singur scale destinate măsurării constructelor respective. Mai mult, literatura domeniului face adesea referiri la aceste caracteristici, considerându-se implicit valabil că psihologul este avizat în legătură cu aspectele tehnice implicate de construcţia şi dezvoltarea scalelor psihodiagnostice.

Pe de altă parte, deşi se recunoaşte tacit că statistica este printre cele mai importante instrumente al profesiunii de psiholog, ea determină încă o reacţie de rezistenţă, sau chiar de respingere din partea studenţilor sau a novicilor, din cauza caracterului ei abstract şi formalizat, sau a evoluţiei sale foarte rapide, prin care demersul matematic devine tot mai sofisticat, parcă anume rezervat unei elite capabile să ţină pasul cu ceea ce s-a numit „imperialismul metodologic" al ştiinţei. Capitolul pe care îl deschidem prin aceste rânduri indică la modul clar cum o cunoaştere psihologică abstractă, de tip statistico-matematic, devine extrem de utilă în contextul psihologic foarte concret al construcţiei de scale. El este destinat a oferi reperele pe care psihologul, pedagogul sau sociologul trebuie să le aibă în dezvoltarea scalelor de măsură a unor constructe specifice domeniului de activitate propriu.

10.2. Paşii de urmat în dezvoltarea scalelor

10.2.1. Pasul întâi: determinăm cu claritate ceea ce dorim să măsurăm

a. Recursul la o teorie valabilă

Ne putem imagina efortul psihologului de a asambla un set semnificativ de mare de itemi presupuşi a avea legătură cu constructul măsurat, de a-l aplica unei populaţii destul de numeroase, de a introduce item cu item datele într-un program de prelucrare statistică, de a

desfăşura o analiză factorială exploratorie pentru a determina care itemi bat în direcţia constructului presupus a fi măsurat şi care în alte direcţii, de a-i selecta şi reuni pe cei mai reprezentativi într-o scală uni- sau multidimensională, de a determina gradul de consistenţă internă (de omogenitate) al fiecărei subscale, stabilitatea ei în timp sau legătura cu constructul măsurat prin corelare cu alte scale presupuse a măsura aceeaşi dimensiune. Un astfel de demers – tipic pentru dezvoltarea unei scale – presupune un mare consum de energie şi el ar putea să nu fie răsplătit cu rezultate corespunzătoare. Aceasta deoarece specialistul trebuie să aibă o cunoaştere şi o conceptualizare prealabilă a domeniului care urmează să fie investigat, fie din parcurgerea literaturii şi a teoriilor în legătură cu problema/ dimensiunea respectivă, fie dintr-o experienţă prealabilă, fie – şi este cazul cel mai fericit – din amândouă direcţiile. Prin aceasta dorim să afirmăm că aspectele tehnice care vor fi invocate în paginile care urmează nu trebuie supraestimate în detrimentul înţelegerii naturii reale a conceptului sau faptului psihologic care urmează a fi măsurat.

Recursul la teoriile cele mai consistente referitoare la acest construct este fundamentală pentru că aceasta ne ajută şi ne ghidează în selectarea elementelor sale esenţiale, care vor fi ulterior transformate în itemii scalei sau ai chestionarului. Multe dintre dimensiunile şi constructele care definesc aspecte psihice, comportamentale sau sociale sunt subtile, ambigue şi nu îşi găsesc o expresie directă sau deplină în experienţa de practician. De aceea este foarte important să ne facem o idee cât mai clară despre ceea ce urmează a fi măsurat. În acest sens teoria este de cel mai mare ajutor, pentru că ea ne ajută să ne facem o idee clară asupra aspectului ce va fi abordat psihometric. Coroborând mai multe puncte de vedere teoretice despre o problemă se degajă elementele sale esenţiale, dar şi neclarităţile sau ambiguităţile, fapt care ajută specialistul să decidă dacă este necesară o nouă direcţie de abordare şi să construiască o primă schiţă a acesteia.

De exemplu, suntem preocupaţi de o problemă practică, cum ar fi cea a tulburărilor de citit-scris, şi observăm că această este mult mai frecventă la băieţi decât la fete. Am remarcat de asemenea faptul că printre disgrafici şi dislexici proporţia stângacilor este de câteva ori mai mare decât în restul populaţiei, ceea ce ne îndreptăţeşte să credem că aceste tulburări sunt asociate cu probleme de lateralizare (manuală şi/sau corticală) a unor funcţii importante. De aceea am dori să creăm un instrument care să fie utilizat ca adjuvant în tulburările instrumentale şi care să diagnosticheze gradul de lateralizare manuală.

Literatura domeniului dominanţei manuale ne atrage atenţia asupra câtorva elemente esenţiale:

- stângăcia poate fi ereditară sau non-ereditară, adică produsă prin alte mecanisme, cum ar fi excesul de testosteron din primele luni de sarcină, care inhibă dezvoltarea emisferei stângi, cel mai adesea dominantă în populaţia de bază, sau stângăcia accidentală, prin scoaterea din uz a mâinii dominante un interval mai lung de 8 luni;
- mecanismul asimetriei cerebrale are o transmitere genetică, gena LRRTM1 având o mare contribuţie la aceasta, ea fiind însă implicată şi în producerea schizofreniei;
- există o genă (*rs* după expresia lui Annett, 1972[17]) care controlează orientarea levogiră sau dextrogiră a unei persoane;
- lateralizarea manuală se asociază strâns cu lateralizarea corticală a limbajului şi cu dominanţa cerebrală;
- lateralizarea manuală este impusă şi de presiunile socio-culturale enorme pentru utilizarea mâinii drepte, presiuni evidenţiabile la aproape toate popoarele, din toate timpurile, în unele culturi ea având şi importante conotaţii religioase;
- rata stângacilor este aproximativ constantă (10% din populaţie) de-a lungul timpurilor, deşi de-a lungul vieţii nuei singure persoane ea pare a fi într-o descreştere accentuată spre vârstele adulte, datorită mortalităţii mai accentuate a stângacilor, de unde şi speranţa de viaţă mai redusă a acestora;
- bolile autoimune, alergiile, tulburările de somn, depresiile şi tentativele de suicid, alături de problemele de învăţat, balbism şi enurezis nocturn sunt mult mai frecvente pentru stângaci;
- problema lateralizării nu se pune doar pentru utilizarea preferenţială a unei mâini, ci şi pentru picior, hemicorp sau pentru organele de simţ cele mai performante (ochi şi urechi), putându-se vorbi de o lateralizare omogenă sau eterogenă (încrucişată);
- problemele de citit-scris-vorbit apar mai frevent la anumite categorii de stângaci (lateralizare încrucişată, nefamilială, patologică etc.);
- s-a avansat ideea că modul de tratare cerebrală a informaţiei de către stângaci este mai curând integrat-holistic şi nu analitic-discursiv ca la dreptaci, de unde şi plusul lor de creativitate;
- stângacii se întâlnesc mult mai frecvent printre cei care urmează meserii vizuale, şi foarte rar în cele dependente de limbă;

[17] Annett, M. (1992). Spatial ability in subgroups of left- and right-handers. In *British Journal of Psyuchology*, 83, pp. 493-515.

- măsurarea gradului de lateralizare cunoaşte câteva metode[18] de tip chestionar, dar pot fi gândite şi metode bazate pe acţiune şi observare (Annett, 1972);
- dincolo de stângăcia „tare" sau dreptăcia „tare", există variantele lor mai moi, o multitudine de probleme fiind asociate mai curând cu lipsa unei lateralizări ferme, decât cu stângăcia în sine;
- stângăcia pare să aibă efecte dezadaptative şcolar mult mai mari printre elevii de gen masculin decât printre cei de gen feminin.

Din toată această masă de informaţii şi din altele de acelaşi tip trebuie extrasă cea de interes pentru scopurile scalei ce va fi construită şi apoi conceptualizată. Operaţionalizarea enormei mase de informaţii despre stângaşi şi stângăcie înseamnă traducerea acesteia într-o formă care să poată fi măsurată, în cazul nostru sub forma itemilor unui chestionar sau a unor probe practice care vor compune itemii unui test de lateralitate. Aşa cum se observă, unele dintre informaţiile de mai sus par a fi mai importante decât altele, având rolul de a ghida întregul demers de construcţie a scalei. Mergând de exemplu pe modelul genetic al lui Annett vom putea determina opt categorii referitoare la lateralitatea manuală. Dacă avem în vedere problema omogenităţii/eterogenităţii lateralizătii va trebui să generăm şi să includem în scală itemi legaţi de preferinţa pentru membrele inferioare, pentru ochi, ureche, hemicorp. Dacă luăm în calcul dimensiunea ereditară a stângăciei vom avea itemi legaţi de frecvenţa acestei particularităţi printre rudele apropiate, iar dacă avem în vedere dimensiunea ei evolutivă pe scala vârstelor vom avea itemi care să distingă între manifestările timpurii şi cele ulterioare ale stângăciei.

b. Nivelul de specificitate al scalei

Din exemplul de mai sus se poate vedea că problema lateralizării poate fi abordată într-un cadru mai larg şi mai general sau într-un cadru mai restrâns şi mai specific. Faptul de a fi mai globală poate fi un avantaj, dar preţul plătit este o lungime mai mare a scalei şi o consistenţă internă mai scăzută. Se ştie că, cu cât nivelul de specificitate este mai ridicat, itemii se corelează mai strâns între ei şi cu constructul sau variabila latentă măsurată. De exemplu, dacă chestionarul de lateralitate construit este unul antropologic, el va include itemi care privesc atitudinea culturală faţă de dextralitate, cum ar fi dacă ai prefera să fii operat de

[18] Oldfield, 1971: *Edinburgh Handedness Questionnaire*; Annett, 1970: *Laterality tests and interests*.

un chirurg dreptaci sau stângaci (Holder, 1992, *Hand Preference Questionnaire*[19]); dacă scala este destinată vârstei de peste 14 ani, ea poate fi una care doar întreabă oamenii despre mâna cu care scriu, desenează, taie, aruncă, se perie, mătură, deschid o cutie etc. (*Edinburgh Handedness Inventory,* EHI, Oldfield, 1971[20]); dacă este destinată determinării obiective a mâinii dominante, se vor folosi itemi acţionali, grupaţi într-o scală ca cea a lui Annett, unde se folosesc în acţiune efectivă cocoloaşe de hârtie, foarfece, chibrituri, periuţă de dinţi, ciocan de jucărie etc. Copiii pot fi examinaţi şi în clasă, cerându-li-se să se joace în perechi pentru a exersa acţiunile care vor fi observate, fapt ce va face posibilă completarea grilei de observaţie sau a chestionarului acţional de către observator.

Într-o abordare de tip obiectiv noi[21] am propus determinarea unui indice de lateralitate prin două probe de tip acţiune, cronometrate: tapping (puncte bătute rapid pe o foaie de hârtie cu creionul, cu mâna dreaptă, apoi cu mâna stângă, câte 10 secunde pentru fiecare, singura constrângere fiind numărul de puncte) şi trasaj (linii verticale, drepte, apropiate, executate rapid, câte 6 secunde cu mâna dreaptă, apoi cu cea stângă, după o execuţie demonstrativă prealabilă a examinatorului).

Gradele de specificitate şi de obiectivitate cele mai ridicate sunt în ultimul exemplu, deoarece probele sunt simple, clare, uşor de testat, iar dispozitivul de înregistrare este unul precis. Mai mult, se operaţionalizează bine problema lateralizării pentru sarcinile motrice grosiere şi pentru cele de fineţe, ambele fiind agregate într-un concludent *Indice motric*.

De această cerinţă a specificităţii se leagă şi problema a ce includem în scală pentru a fi măsurat. În cazul de mai sus distincţia este evidentă: itemii sunt de tip autoraportare (primul exemplu), rezultaţi din observaţie (al doilea exemplu) sau rezultaţi din măsurarea efectivă a unor comportamente induse (al treilea exemplu).

Dacă am alcătui o scală care amestecă cele trei maniere de obţinere a itemilor ar rezulta probabil un instrument compozit, eterogen, mai puţin maniabil. Problema se referă însă nu numai la cum sunt produşi itemii, ci şi la conţinutul lor. De exemplu, dacă concepem anxietatea ca pe un fenomen psihic (o teamă difuză, fără obiect) nu are rost să amestecăm elemente somatice într-o fenomenologie psiho-comportamentală. Dar dacă am relevat conotaţiile psihosomatice ale acesteia, putem concepe şi operaţionaliza anxietatea mult mai nuanţat, dintr-o parte afectivă, una cognitivă şi una somatică, cum a procedat Burns în

[19] Holder, M.K. (1992). *Hand Preference Questionnaire: One Gets What one Asks For*. M. Phil thesis, Department of Antropology, Rutgers University, New Jersey, USA.
[20] Oldfield, R.C. (1971). The assessment and analysis of hanhdedness: The Edinburgh Inventory. *Neuropsychologia*, 9, pp. 77-114.

chestionarul său. Acesta şi-a construit *Inventarul de anxietate* astfel încât indicele global obţinut să fie o expresie a sentimentelor anxioase, a gândurilor anxioase şi a simptomelor anxioase de ordin fiziologic. Această perspectivă asupra anxietăţii este mult mai nuanţată şi mai utilă într-o cercetare despre stres decât una care ar evalua doar aspectele cognitive sau afective ale acesteia.

Pe de altă parte, într-un chestionar de scheme cognitive anxioase ar trebui respectată reţeta cognitivistă de abordare, fiind eliminate referirile somatice la anxietate, în favoarea cogniţiilor patologice.

10.2.2. Pasul al doilea: generarea unui set mare de itemi

a. Lungimea şi redundanţa setului iniţial de itemi

Pasul acesta presupune crearea efectivă a unui mare număr de itemi, posibili candidaţi la forma finală a scalei. Aceştia trebuie să fie gândiţi astfel încât să fie efectiv în legătură cu constructul latent măsurat, deşi dovezile în acest sens vor fi culese mai târziu. Scopul de ansamblu al scalei va fi cel care ne ghidează în procesul de generare al itemilor, care sunt proiectaţi fiecare ca un mic test separat, corelat cu variabila latentă măsurată. Dacă fiecare item este sensibil asociat cu această variabilă, prin agregarea unui număr mare de itemi se speră ca scala rezultată să fie o expresie mult mai concludentă a variabilei sau a constructului respectiv. Un test/scală bun(ă) este aşa datorită celor mai mici unităţi care îl compun şi care sunt itemii săi.

Universul itemilor care descriu un construct psihologic este foarte larg şi deschis, de aceea nu se pot stabili cu precizie graniţele sale. Vom fi de acord însă că putem asambla un număr mare de itemi, că aceştia pot corela ridicat între ei la un nivel ridicat, scala fiind una omogenă, fără ca doar prin aceasta ea să fi prins esenţa constructului respectiv. Fericire, sens existenţial, satisfacţie în muncă, adaptare şcolară sau la locul de muncă, complianţă, identitate sexuală, satisfacţie maritală etc. sunt concepte vagi, care au nevoie de un cadru teoretic care să le clarifice, dar şi de o bună experienţă pentru a genera itemii care să le surprindă esenţa.

O problemă importantă a acestei faze de construcţie a scalei este cea a **redundanţei**: trebuie ea evitată, trebuie admisă, şi până la ce nivel? Teoreticienii admit că, atunci când vrem să construim o scală, redundanţa nu este deloc un lucru rău, aceasta deoarece itemii redundanţi sunt şi intercorelaţi, prezenţa lor ducând la creşterea consistenţei interne a scalei.

---

[21] Clinciu, A.I. (2003). *Vizual şi auditiv în structura cognitivă umană*. Teză de doctorat nepublicată. Universitatea Bucureşti, Facultatea de Psihologie şi Ştiinţele Educaţiei.

Singura problemă este să avem o redundanţă bună, folositoare, care se referă la constructul măsurat, şi nu la aspecte incidentale ale itemilor, cum ar fi acelea legate de detalii minore de formulare. Când exprimi aceeaşi idee în două moduri distinct diferite redundanţa respectivă poate fi utilă, pentru că nu putem şti cu suguranţă care dintre formulări este mai diagnostică şi mai bine acceptată de subiecţi. De exemplu „Nu am nici un fel de jenă în a aborda persoane necunoscute" şi „Îmi vine uşor să intru în vorbă cu persoane necunoscute" sunt enunţuri foarte similare, deoarece au în vedere dimensiunea inhibiţiei, definitorie pentru fobia socială. Diferenţele nu sunt însă pe sensul principal al enunţurilor, ci pe conotaţiile rezultate din modul de exprimare al ideii: în prima situaţie propoziţia începe negativ, în a doua afirmativ; în prima cuvântul *jenă* are şi conotaţii etice („fără ruşine"), ceea ce nu este cazul în a doua; în prima formulare *a aborda* este mai pretenţios decât mai familiarul *a intra în vorbă*. Nu ştim însă care dintre cele două formulări este cea mai bună şi de aceea le putem include în forma iniţială a scalei, urmând să amânăm decizia legată de selecţia celui mai bun item până la prelucrarea datelor rezultate din aplicarea instrumentului.

O regulă spune că putem tolera o oarecare redundanţă şi în faza finală de construcţie a unei scale, deşi teoretic este permisă mai multă redundanţă în faza iniţială decât în cea finală, pentru a da ocazia constructorului scalei să-şi manifeste preferinţele şi să selecteze forma mai bună a itemilor redundanţi.

Problema redundanţei angajează direct o problemă la fel de importantă, cea a **numărului de itemi** care compun setul iniţial. Noi ştim că un atribut fundamental al scalei este consistenţa sa internă şi că aceasta este dependentă de două lucruri: de gradul de corelare al fiecărui item cu scorul final la constructul măsurat şi de numărul de itemi ai scalei. De aceea putem afirma că, cu cât un set iniţial de itemi este mai larg, cu atât este mai bine, deoarece avem de unde selecta. Trebuie ştiut însă foarte bine că un set de itemi prea extins, dintre care mulţi se repetă, poate genera consumuri de timp suplimentare, sau chiar plictiseală, putând compromite sinceritatea şi autenticitatea răspunsurilor. De aceea cercetătorul poate elimina din start itemii neclari, cu relevanţă scăzută sau cu prea multă similaritate, pentru a rămâne în final la un set iniţial rezonabil, pe care să-l aplice într-o singură şedinţă, pentru a nu avea prea multă „moarte experimentală".

Deşi nu există reguli fixe, paritatea dintre lotul iniţial de itemi şi cel final poate fi de de 4/1 sau 3/1 (30-40 de itemi iniţiali pentru o scală de 10 itemi) sau, cel mai adesea de 2/1, atunci când itemii sunt construiţi cu grijă şi cu o bună cunoaştere a constructului măsurat. Deoarece forma iniţială depinde de cea finală, cât de **lungă** trebuie să fie o scală pentru a avea calităţi psihometrice bune? Este un răspuns dificil de furnizat, deoarece fiecare item

poate fi considerat ca o singură scală şi există chiar scale alcătuite dintr-un singur item.[22] În mod obişnuit, cu cât o scală are mai mulţi itemi, cu atât mai mult ea poate dispune de calităţi psihometrice (fidelitate, validitate, sensibilitate) mai bune; 50-60 de itemi pentru o scală unidimensională reprezintă limita de sus, iar 10-20 de itemi limita de jos. Scalele sub 10 itemi pot fi utile instrumente de *screening*, dar nu diferenţiază decât grosier între subiecţi, pe când scalele lungi dau diferenţieri de mai mare fineţe.

b. Reguli de scriere a itemilor unei scale

Deşi procesul cel mai delicat, mai creativ şi interesant din toată ciclicitatea descrisă aici, scrierea itemilor nu este strict normată, lăsând loc masiv inspiraţiei, creativităţii şi experienţei în domeniu. Scopul acestei etape este pur şi simplu acela de a identifica o multitudine de maniere în care constructul se leagă de viaţa reală sau de comportament. Pentru aceasta recursul la instrumente similare, la literatura (adnotată) a problemei pot fi foarte utile („itemi de import"), dar la ce bun să încerci să compui o scală când deja există altele similare? Motivul pentru care încercăm altceva este pentru că sperăm să o facem mai bine decât alţii sau pentru că mergem pe o cale nouă şi nebătătorită. A scrie repede, la inspiraţie şi relaxând spiritul critic poate fi cel mai recomandat, deşi personal cred că această fază este cu atât mai productivă cu cât creativitatea şi spontaneitatea ei sunt mai bine pregătite prin faza de documentare teoretică (parcurgerea literaturii aferente pentru a ne face o idee cuprinzătoare şi nuanţată despre constructul în cauză) şi prin faza de incubaţie/gestaţie, când lăsăm informaţia „să dospească" şi să se decanteze.

Un creator avizat ştie că înainte de a se aşterne pe scris itemi el trebuie să aibă o reprezentarea complexă şi nuanţată a constructului de măsurat; că a notat determinările esenţiale ale acestuia pe o foaie de hârtie, pe care le-a studiat încercând să le pună mental în legătură cu tema scalei sale, pentru a-l orienta ca un fel de ghid. De exemplu, fericirea conjugală depinde ea doar de calitatea vieţii sexuale a cuplului, sau şi de implicarea soţilor în responsabilităţile casnice (creşterea copiilor, elaborarea bugetului şi a modului de cheltuire a

[22] *Torque Test* este o probă cu un singur item care îşi propune să determine simultan preferinţa manuală şi dominanţa cerebrală. Subiectul îşi scrie numele cu o mână şi apoi îl încercuieşte, după care face acelaşi lucru cu cealaltă mână. Mâna care produce cea mai frumoasă scriere este cea preferată; dacă ambele cercuri sunt orientate în sensul acelor de ceasornic, dominanţa cerebrală este de dreapta; dacă ambele cercuri sunt orientate în sensul invers acelor de ceasornic, dominanţa cerebrală este de stânga; dominanţa cerebrală este mixtă în situaţia în care un cerc are o orientare şi celălalt orientarea inversă.

Testul pare seducător prin simplitate şi acurateţe, dar are multe inconveniente. Asfel, mai ales în cazul scrierii, preferinţa manuală poate fi impusă prin educaţie, subiectul efectuând multe alte acţiuni cu cealaltă mână. La o singură aplicare poate interveni cu o oarecare probabilitate hazardul; nu ştim nimic de faptul dacă preferinţa pentru mână se extinde şi asupra piciorului etc. (Vezi Molly Kalafut, *Left-handedness. Nature or Nurture, Blessing or Course?).*

banilor, implicarea în curăţenie), de calitatea comunicării dintre soţi, de gradul de independenţă reciprocă acordat, de valorile comune, de toleranţa faţă de profesia celuilalt, de acceptarea prietenilor partenerului, de multitudinea de proiecte şi interese comune, de jocurile comune, de aspectele hedonice şi nonhedonice ale relaţiei etc.?

Deşi este foarte greu de fixat reguli, dăm mai jos câteva caracteristici ale **itemilor buni şi răi**, indicând şi cele mai frecvente erori care pot apărea în scrierea acestora.

- Deşi nu este indicat să sacrificăm înţelesul itemilor în favoarea scurtimii lor se recomandă ca itemii să fie **scurţi şi clari**. Putem formula acelaşi item astfel: „La o petrecere reţin cu uşurinţă chipurile şi numele musafirilor"; „Îmi trebuie destul de mult timp şi am adesea mari probleme în a memora de la prima întâlnire numele sau chipurile invitaţilor la o petrecere mai mare". Al doilea enunţ este de evitat pentru că este lung şi neclar, pentru că introduce termenii *destul de mult* şi *mai mare* care sunt vagi.
- **Nivelul de dificultate al citirii** la care sunt scrişi itemii este de asemenea important. Iată doi itemi din *Miller Marital Locus of Control*: „Adesea găsesc comportamentul soţului meu nepredictibil"; „Când suntem în conflict, soţul meu este acela care reconciliază cel mai adesea situaţia". Cuvintele *nepredictibil* şi *reconciliază* ţin de un nivel mai ridicat de cultură, fiind pretenţioase în raport cu omul comun. Deşi există o mulţime de metode de a determina vârsta lecturii, s-a căzut de acord că nivelul de lectură al omului care citeşte ziarul este corespunzător clasei a şasea (12 -13 ani). Multe chestionare de personalitate (MMPI, CPI) presupun o vârstă a lecturii tipică acestui nivel, fenomen de care dacă nu ţinem seama, reducem spectrul de utilizare al unui instrument doar la anumite categorii de persoane. O propoziţie de lungime medie pentru clasa a şasea are între 14 - 18 cuvinte şi 24 de silabe; pentru clasa a şaptea ea are 18 cuvinte şi 24 de silabe. Pentru majoritatea instrumentelor care aspiră spre o utilizare extensivă în populaţie nivelul de lectură anticipat este între clasa a cincea şi a şasea. Dificultatea citirii şi înţelegerii unui text provine nu numai din lungimea sa, ci şi din caracteristicile sintactice şi semantice ale itemului.
- Cele mai frecvente **erori de evitat** sunt negaţia, dubla negaţie şi enunţurile dublu încapsulate (cu dublă determinare). De exemplu formularea „Nu-mi place să nu-mi fac temele" trebuie evitată, pentru că răspunsul solicită un efort de analiză mare iar rezultatul final este nesigur. De fapt şi negaţia simplă creează probleme de claritate: la itemul „Nu-mi plac rromii" răspunsul Nu poate să semnifice simultan acordul (nu, nu-

mi plac) sau dezacordul (nu sunt de acord cu propoziţia respectivă). Dubla determinare este mai greu de sesizat. „Mă enervează la culme invidia şi dorinţa de răzbunare a oamenilor" pune respondentul în dificultate atunci când sursa enervării sale este ori numai invidia, ori numai dorinţa de răzbunare. Scorul pozitiv la acest item nu distinge dacă prima, a doua sau ambele determinări trebuie să fie prezente pentru a răspunde afirmativ la itemul respectiv. Alt exemplu: „Sprijin drepturile rromilor pentru că discriminarea este un păcat pedepsit de Dumnezeu". Pot să sprijin drepturile rromilor pentru că discriminarea este împotriva valorilor democraţiei, pe care eu le apăr, şi nu pentru ea că este un păcat pedepsit de instanţa divină.

- Trebuie evitate pe cât se poate erorile de formulare care produc **ambiguitate**. „Reprezentanţii noştri din guvern ar trebui să susţină cu mai mulţi bani cercetarea ştiinţifică" induce confuzie din cauza pronumelui *noştri*: se referă la noi ca popor, la noi ca grup de partid sau la noi, cei din sistemul de învăţământ? Alteori confuzia pleacă de la utilizarea formei adjectivale în locul substantivului: „Tratamentul diabetului trebuie să fie o prioritate a sistemului de sănătate" sună mai bine şi mai clar decât „Tratamentul diabeticilor trebuie să fie o prioritate a sistemului de sănătate", pentru că un sistem de sănătate se luptă cu boala la modul cel mai general, şi abia în secundar cu cei care i-au căzut deja victimă. La aceste surse de ambiguitate se mai adaugă şi propoziţiile cu dublu înţeles. „Guvernul N. a dat în primire corect gestiunea ţării" poate produce multă ilaritate şi confuzie printre respondenţi din cauza sintagmei *a da în primire* care în jargon înseamnă a muri.
- **Propoziţii cu conţinut pozitiv şi negativ** prezente în aceeaşi scală constituie o precauţie necesară pentru a evita distorsionarea răspunsurilor subiectului prin nevoia de conformism, de a fi de acord sau de a consimţi. În chestionarul de *Stimă de sine* (RSE) al lui Rosenberg (1965) propoziţia „Aş vrea să am mai mult respect pentru mine" este urmată de „Uneori mă simt inutil", ambele afirmative ca formă, dar una cu conţinut pozitiv iar celaltă negativ. Acest lucru obligă respondentul să evalueze cu atenţie fiecare item înainte de a da un răspuns, ceea ce duce la evitarea tendinţelor de contaminare pe care l-ar produce formularea exclusiv pozitivă a itemilor. DeVellis şi Callahan (1993)[23] remarcă faptul că această precauţiune de a introduce itemi cu polaritate inversă, negativă, are un preţ, deoarece se poate produce o confuzie (mai

[23] DeVellis, R.F., Callahan, L.F. (1993). A brief measure of helplesness: The helplesness subscale of Rheumatology Attitude Index. *Journal of Rheumatology*, 20, pp. 866-869.

ales în cazul chestionarelor lungi) între gradul de acord al respondentului şi conţinutul itemului, fără a mai ţine cont de polaritatea negativă şi de exprimarea tăriei atributului de măsurat. Concluzia ar fi că itemii formulaţi în direcţie opusă au performanţe mai slabe decât cei cu orientare directă şi că dezavantajele includerii lor în scală atârnă mai greu decât presupusele beneficii. Soluţia ar fi, mai ales pentru chestionarele mai lungi, ca numărul acestor itemi să fie mai mic decât cei formulaţi pozitiv, dar să fie diseminaţi printre aceştia.

10.2.3. Pasul al treilea: determinarea tipului de format al măsurătorii

Această chestiune are în vedere mai puţin itemul şi mai mult formatul tipului de scală pe care subiectul va răspunde. Deşi mulţi autori nu fac distincţia între o scală şi un index aceasta există şi trebuie luată în considerare. O scală diferă de un index prin aceea că într-un index itemii nu au nevoie să fie puşi într-o anumită ordine, fiecare item având aceeaşi pondere şi importanţă, ceea ce permite însumarea lor cu scoruri egale, rezultatul fiind chiar indexul. În sens foarte larg, o scală este un set de numere ori de alte simboluri utilizate pentru a desemna caracteristici ale unei variabile şi care este utilizată pentru măsurare. Numerele de pe un termometru sau cuvintele „mic" „mediu" şi „mare" de pe cutia de aer condiţionat sunt scale. Vogt (1999, p. 256) distinge şi un înţeles mai special al cuvântului scală care este „un grup de măsuri relaţionate cu o variabilă. Itemii dintr-o scală sunt aranjaţi într-o anumită ordine a intensităţii sau a importanţei lor."[24]

a. Scalele Likert sau de tip Likert

Sunt cele mai larg utilizate în ştiinţele sociale pentru determinarea atitudinilor datorită faptului că sunt uşor de construit şi tind să aibă o mare fidelitate, ele fiind potrivite şi pentru cosntrucute multidimensionale. Răspunsurile respondentului sunt gradate de la *Puternic de acord, De acord, Nici acord, nici dezacord, Dezacord, Dezacord puternic*. Tipul de expresii verbale pentru a măsura intensitatea poate varia considerabil, subiecţii putând fi chestionaţi dacă *aprobă în totalitate, aprobă parţial, dezaprobă uşor, dezaprobă în totalitate*. Deoarece scalele Likert sunt destinate opiniilor, credinţelor şi atitudinilor, este o problemă alegerea răspunsurilor potrivite pentru categoriile extreme, care nu trebuie să fie de genul *întotdeauna* sau *niciodată,* pentru că aceste situaţii se întâlnesc destul de rar în viaţa reală. În consecinţă

[24] Vogt, W.P. (1999). *Dictionary of Statistics and Methodology. A Nontechnical Guide for the Social Sciences*. Sec. edition. Thousand Oaks, London, New Delhi: Sage Publications, p. 256.

ele vor fi alese cât mai rar, pentru a nu contribui prin aceasta la descreşterea variabilităţii scorurilor pentru itemul respectiv.

De fapt, scalele Likert au o interesantă caracteristică, rezultată din faptul că ele au două modalităţi de a grada intensitatea constructului respectiv pentru populaţia studiată. Una ţine chiar de modalităţile de gradare a acordului/dezacordului din variantele de răspuns propuse, fapt de care am vorbit deja.

O a doua modalitate ţine de tăria/slăbiciunea modului cum este exprimat itemul însuşi. De exemplu, opinia despre politicienii români post-decembrişti ar putea fi exprimată la trei niveluri de tărie („duritate”):

1. „Dezastrul moral şi economic din România post-revoluţionară este integral opera politicienilor incompetenţi pe care i-am avut.”
2. „Deoarece nu am avut o clasă politică constituită în timp, liderii politici ai României post-revoluţionare au întârziat evoluţia economico-socială a ţării din cauza greşelilor provocate de lipsa lor de experienţă.”
3. „În pofida unor greşeli cauzate de lipsa de experienţă, politicienii români sunt pe cale să construiască o democraţie reală în România post-revoluţionară.”

Cum tendinţa obişnuită a majorităţii oamenilor este aceea de nu fi – din bun simţ – prea ofensiv-agresivi, prima afirmaţie ar putea discrimina doar printre cazurile extreme, deoarece este una foarte tare. Ultima afirmaţie, care este cea mai moale şi mai împăciuitoristă are un caracter mai agreabil, ar putea conveni majorităţii oamenilor, dar ar putea indica mai mult lipsa opiniei, decât prezenţa ei. Ori o cercetare caută mai curând ceea ce este şi se manifestă (prezenţa), decât ceea ce nu este (absenţa). Aşadar este o iluzie să credem că un item poate baleia întreg spectrul unei opinii în unităţi de lungime egală. Avem doar posibilitatea ca, din combinarea tăriei/slăbiciunii modului de formulare a itemilor şi din tipurile de gradaţii de pe scala de răspuns folosită să „centrăm” itemul pentru zona populaţională şi pe problema care ne interesează. Itemul trebuie să răspundă la întrebarea: „Cum sunt dispuşi să răspundă oamenii la diferitele intensităţi sau tării ale constructului măsurat prin acest item?”.

Prezentăm mai jos două exemple de itemi construiţi pe scala Likert.

Prin sistemul reformei de la Bologna, învăţământul românesc s-a integrat bine învăţământului european.

1. Dezacord puternic	*2. Dezacord moderat*	*3. Dezacord uşor*	*4. Acord uşor*	*5. Acord moderat*	*6. Acord puternic*

Combaterea consumului de ţigări este o prioritate naţională de vârf.

1. Complet adevărat	*2. Mai curând adevărat*	*3. Aşa şi aşa*	*4. Mai curând neadevărat*	*5. Complet neadevărat*

b. Scalele Thurstone

Modalitate complexă, rafinată şi dificilă de a construi itemii unei scale, iniţiată de Thurstone, acest tip de scală şi-a pierdut mult din relevanţă în favoarea scalelor Likert. În principiu eleganţa şi dificultatea metodei vine din pre-calibrarea itemilor şi asamblarea lor pentru a răspunde la diferite niveluri de prezenţă în item a atributului sau a constructului măsurat. De regulă, în acest tip de scală răspunsurile nu mai au o multitudine de nuanţe, deoarece gradarea intensităţii este pe verticală (mai mulţi itemi asamblaţi pe o aceeaşi coloană în funcţie de mărimea progresivă a atributului măsurat), şi nu pe orizontală (modalităţi din ce în ce mai accentuate ale intensităţii atributului la un singur item, tipic pentru scala Likert). Astfel, pentru aceeaşi idee se generează mai mulţi itemi (care au formulări diferite) care să reprezinte intervale egale de-a lungul unui continuum vertical, ceea ce conferă scalei proprietăţi matematice remarcabile, pentru că se pretează la tratamente statistice tipice scalelor de intervale. Nunaly (1978) susţine că este incomparabil mai uşor să explici principul unei scale Thurstone decât să o construieşti. A găsi itemi care să „rezoneze" la intervale egale cu constructul măsurat (ca un diapazon acordat pe o anumită frecvenţă) este destul de dificil. Acest deziderat exprimat de scalele Thurstone şi-a găsit totuşi o manieră de rezolvare prin *Teoria răspunsului la itemi*, unde calibrarea acestora urmăreşte totuşi un cu totul alt demers.

Prezentăm mai jos două răspunsuri tipice pentru scala Thurstone.

Exemplul 1.

Când un elev are dificultăţi de învăţare, profesorii:

	Adevărat	Fals
Îl insultă		
Îl critică sau ironizează		
Îi întrerup răspunsul		
Nu au nici o reacţie		
Îi explică suplimentar		
Îl încurajează		
Fac cu el pregătire după ore		

Exemplul 2.

a. Faptul de a termina cu medie mare şcoala şi de a prinde un liceu bun este tot ce-mi doresc ca răsplată de la copilul meu. Acord ..... Dezacord .....

b. Pentru mine ca părinte cea mai mare răsplată din partea copilului este să fie sănătos şi să meargă la şcoală. Acord ..... Dezacord .....

c. Fericirea copilului meu nu are nimic de-a face cu ceea ce el dobândeşte prin educaţia şcolară sau cu scopurile materiale. Acord ..... Dezacord .....

Modalitatea de gradare pe verticală şi de construire a ierarhiei nu este o operaţiune foarte simplă. Metoda utilizată de Thurstone (1929) a fost cea a comparării perechilor de răspunsuri, deoarece ea presupune ca subiecţii-observatori să judece de exemplu enunţurile A, B, C luate câte două (A cu B, A cu C şi B cu C), să evalueze din frecvenţa aprecierilor ordinea ierarhică a celor trei enunţuri, după care să construiască un indice pentru distanţa dintre ele, care va fi exprimat în termeni de probabilitate prin raportarea la curba lui Gauss. Scala obţinută în felul acesta va putea fi verificată: pornind de la valorile ei trebuie să se poată face predicţii asupra viitoarelor alegeri ale indivizilor.[25]

c. Scalele Guttman

Acestea sunt un tip de scală creat de Louis Guttman. Original ea a fost destinată să fie utilizată după ce datele au fost colectate, pentru a vedea dacă ietemii dintr-un index pot fi aranjaţi într-o scală în funcţie de tăria sau de puterea lor. De exemplu, într-o cercetare despre avorturi itemii unui chestionar pot fi aranjaţi ierarhic ca intensitate astfel:

Aţi fi de acord ca femeile să facă avort dacă:

1. A nu face avort i-ar periclita viaţa proprie?
2. Fătul este malformat?
3. Este prea săracă pentru a se putea ocupa în mod adecvat de copil?
4. Ea nu mai doreşte să aibă copii?

Cine ar răspunde la această întrebare ar face-o într-o manieră scalară: persoanele care au răspuns afirmativ la întrebarea 4 ar fi răspuns afirmativ şi la întrebările 1, 2, 3; cele care au răspuns la întrebarea 3 ar fi răspuns şi la întrebările 1 şi 2, dar nu obligatoriu şi la 4, şi aşa mai departe. Dacă scala nu ar fi de tip Guttman, nu ar exista un astfel de pattern şi răspunsul afirmativ la numărul 4 nu ar implica răspunsurile afirmative şi la celelalte 3 întrebări.

Scalele Guttman sunt foarte utile şi indicate pentru informaţiile obiective sau în situaţiile în care este logic necesar ca, dacă ai răspuns afirmativ la un anumit nivel dintr-o ierahie, aceasta înseamnă să faci acelaşi lucru şi pentru nivelurile inferioare ale ierarhiei. De exemplu, la întrebarea dacă fumaţi, variantele de răspuns pot fi organizate ierarhic astfel: nu fumez deloc; fumez sub 10 ţigări pe zi; fumez sub 20 de ţigări pe zi; fumez sub 30 de ţigări pe zi; fumez mai mult de 30 de ţigări pe zi. Variantele de răspuns sunt aranjate astfel că nivelul la care este prezent atributul este indicat de cel mai de sus item la care subiectul a dat un răspuns afirmativ, cei de sub el fiind şi ei implicit afirmativi (deoarece reprezintă niveluri mai scăzute ale atributului). Deci această scală presupune o gradaţie şi decelarea punctului

[25] Vezi Laurens, S., Moscovici, S. (2007). Construirea scalelor. In S. Moscovici, F. Bruschini (coord.), *Metodologia ştiinţelor socioumane*. Iaşi: Polirom, pp. 482-414.

critic în care răspunsul afirmativ devine negativ. Dacă pentru informaţiile obiective scala Guttman este foarte indicată, pentru alte categorii de fenomene este dificil de ordonat răspunsurile într-o manieră uniformă pentru toţi indivizii, fapt care limitează utilizarea acestui tip de scală. Pentru situaţiile în care ordonarea itemilor este importantă, modelele bazate pe teoria răspunsului la itemi par a oferi soluţii mai adecvate. Prezentă mai jos doi dintre itemii scalei de depresie Beck, asamblaţi pe o scală Guttman.

A. TRISTEŢE

1. Nu mă simt trist.
2. Mă simt trist.
3. Sunt trist tot timpul şi nu pot scăpa de tristeţe.
4. Sunt atât de trist şi de nefericit încât nu mai pot suporta.

B. PESIMISM

1. Viitorul nu mă descurajează.
2. Mă simt descurajat când mă gândesc la viitor.
3. Simt că nu am ce aştepta de la viitor.
4. Simt că viitorul este fără speranţă şi nimic nu se mai poate îndrepta.

d. Scala de distanţă socială Bogardus

Acest tip de scală a apărut în 1923, fiind destinată să măsoare distanţa socială, adică distanţa care separă un grup de altul. De aici rezultă posibilitatea de a da o expresie măsurabilă raporturilor de discriminare socială dintre grupurile etnice, naţionale sau religioase. Scala Bogardus poate fi considerată în acelaşi timp şi o scală de proximitate deoarece indică cea mai mică proximitate acceptabilă pentru un individ faţă de grupul luat în considerare. Individul este chestionat dacă ar fi dispus, de exemplu, să trăiască în aceeaşi ţară cu rromi; să trăiască în aceeaşi localitate cu aceştia; să aibă vecini rromi; să locuiască în aceeaşi casă cu rromi; să aibă prieteni de etnie rromă; să aibă rude rromi; să aibă un copil care se căsătoreşte cu o persoană de etnie rromă.

Populaţiile pot fi comparate simultan trecând pe verticală categoriile de proximitate iar pe orizontală populaţiile ce vor fi comparate: negri, latino, europeni, asiatici, sau turci, evrei, polonezi, spanioli etc., pentru America. Sau români, maghiari, saşi, ţigani, ruteni etc. perntru România. Evident, distanţele studiate sunt unele sociale şi nu fizice, dar exprimarea în termeni fizici a acestora este judicioasă. Acest tip de scală permite compararea atitudinilor grupurilor faţă de alte grupuri deoarece enunţurile despre obiectul social studiat pot să fie astfel clasate pentru a indica fie proximitatea, fie distanţa. „Scalele construite în acest fel dau rezultate în acelaşi timp interesante şi valide” apreciază Laurens şi Moscovici (op. cit., p. 391).

e. Numărul optim al categoriilor de răspuns

Cei mai mulţi dintre itemii scalelor se prezintă sub forma unui trunchi comun care este enunţul de bază al itemului, şi o categorie de descriptori care să indice gradul de acord/dezacord al respondentului cu enunţul respectiv. Atenţia noastră ce centrează acum pe această a doua categorie. Statistica ne învaţă că o caracteristică importantă a unui item sau set de itemi este variabilitatea/varianţa. Teoria răspunsului la itemi determină tocmai relaţia existentă între variabilitatea itemului şi a scalei ca întreg. Cu cât este mai mare această variabilitate cu atât este mai bine, deoarece variabilitatea itemilor contribuie la creşterea variabilităţii de ansamblu a întregii scale. Cu cât ea este mai mare, cu atât se pot determina mai multe categorii de scor şi deci scala este mai discriminativă.

Variabilitatea unei scale depinde de doi factori: de numărul itemilor (cu cât mai mulţi itemi, cu atât mai multă variabilitate) şi de numărul variantelor de răspuns pe care aceştia le prezintă (cu cât mai multe, cu atât mai mare este şi variabilitatea itemului). De aici şi interesul pentru problema abordată în acest paragraf. Vom distinge deci existenţa unor itemi cu o manieră de scorare săracă. Deoarece sunt necesare cel puţin două posibilităţi de răspuns, acest mod de operare poartă numele de **scorare binară**. În măsura în care există posibilitatea de a avea o multitudine de clase de scor, să zicem 100, cum se poate distinge fără falsă precizie între nivelurile 66 şi 67 de anxietate, de exemplu? Cu cât mai multe clasele de răspuns, cu atât posibilitatea de a diferenţia realmente între valorile de scor apropiate descreşte. În practică rareori se folosesc mai mult de 7-11 clase de scor şi foarte frecvent între 3 şi 7.

Opţiunea binară de scorare este frecvent utilizată, căci, deşi au cea mai mică variabilitate a scorurilor per item, scalele de acest tip se completează extrem de facil şi rapid, lăsând deschisă posibilitatea de a introduce un număr mai mare de itemi pentru a creşte variabilitatea de ansamblu a scalei sau pentru a vedea cum noi categorii de fapte coroborează cu constructul măsurat. Această manieră de scorare este tipică – dar nu exclusivă! – scalelor Thurstone şi Guttman datorită complexităţii de construcţie a itemului, care agregă pe verticală mai multe enunţuri gradate ierarhic. De asemenea, acest tip de itemi este numit şi cu alegere forţată[26]. Datorita uşurinţei cu care se lucrează, opţiunea binară este extrem de utilă în

[26] Atenţie, itemii cu alegere forţată nu sunt numai cei binari, de tipul Adevărat/Fals sau Da/Nu, ci şi unii itemi pari pentru care, neexistând varianta ehivocă din mijloc („aşa şi aşa"), subiectul trebuie să opteze clar într-o direcţie sau în alta.

completarea listelor de adjective prin care se investighează trăsăturile de personalitate (modelele Big Five).

Dacă la un pol sunt opţiunile binare, la celalălalt pol este numărul mare sau foarte mare de clase de scor care pot aduce, cum am arătat anterior, o falsă precizie, adică o falsă variabilitate. Din statistică ştim că variabilitatea poate fi descompusă într-o parte bună, numită şi *sistematică,* deoarece este strâns legată de oscilaţiile reale ale prezenţei constructului măsurat în item, şi dintr-o parte rea, de *eroare*. Deoarece creştarea variabilităţii pe scalele cu număr mare de variante de răspuns se face în contul erorii, şi nu al variabilităţii sistematice, nu este indicat să avem prea multe valori de scor. Numărul acestora este impus finalmente de abilitatea respondenţilor de a discrimina semnificativ între variantele de scor propuse pentru fiecare item. Această capacitate depinde şi de aranjarea în pagină a variantelor de răspuns asfel încât acestea să sugereze un continuum. Astfel secvenţa:

Aproape niciodată *Rareori* *Din când în când* *Deseori* *Aproape mereu*

sugerează bine un continuu temporal, divizat în unităţi discrete de tipul frecvenţei de apariţie.

O problemă importantă este dacă optăm pentru un număr impar de categorii de răspuns, ceea ce în principiu este bine, pentru că se creează o anumită simetrie în jurul unui punct zero natural. Acest fapt oferă însă de multe ori indecişilor posibilitatea refugierii în răspunsul evaziv din mijloc. Numărul par sau impar de opţiuni este mai puţin important decât modul cum sunt aranjate variantele de răspuns astfel ca distanţele dintre ele să fie aproximativ egale pentru a exprima gradul de acord sau de dezacord. Aceasta pentru că datele vor fi codificate numeric prin cifre (1, 2, 3, 4, 5 pentru cinci variante de răspuns, de la *foarte puţin* sau deloc, la *foarte mult* sau în întregime, ori invers 7, 6, 5, 4, 3, 2, 1 pentru o scală care începe de exemplu cu *totdeauna* şi se termină cu *niciodată*). Se poate construi astfel un continuum divizat într-un număr de trepte (3-7 trepte), de la dezacordul total, la acordul deplin sau invers.

În legătură cu punctul de simetrie din mijloc, el poate avea semnificaţii diferite. Astfel *nici acord, nici dezacord* indică mai curând o atitudine apatică, pe când *la fel de mult acord şi dezacord* sugerează o atracţie egală faţă de ambele variante de răspuns. Deci în cazul numărului impar de variante de răspuns trebuie să controlăm ca subiectul să nu selecteze varianta neutră doar pentru a evita o alegere, oferind astfel un răspuns echivoc, cu o valoare discutabilă. În majoritatea situaţiilor cercetătorul încearcă să evite echivocul punctului central şi să forţeze o alegere clară, dar trebuie spus că opţiunea pentru variante pare/impare de răspuns este funcţie de întrebare, de răspus şi de scopul investigaţiei. Într-o cercetare cu două

variante extreme de răspuns (ce situaţii preferi: *pe cele plicticoase?*; *pe cele primejdioase?*) introducerea variantei de răspuns *între cele două* poate fi pe deplin justificată, subiecţii neputând alege întotdeauna între situaţii atât de tranşant delimitate.

Ca tip de format al răspunsurilor amintim şi **diferenţialul semantic** al lui Osgood şi Tannenbaum (1955). Utilizat în studierea atitudinilor de grup, acest mod de scalare presupune identificarea unui grup ce este studiat (să zicem avocaţi) şi a unei liste de perechi de adjective.

Onest	__ __ __ __ __ __ __ __ __	Neonest
Competent	__ __ __ __ __ __ __ __ __	Incompetent
Vorbăreţ	__ __ __ __ __ __ __ __ __	Taciturn

Liniuţele (de regulă 7 sau 9) reconstituie un continuu între polii definiţi de cele două adjective, subiectul marcând cu un X punctul de pe acest continuu pe care el crede că se află membrii categoriei respective, după care trece la ratingul următoarei perechi de adjective, până la epuizarea întregii liste. Prin prelucrarea separată pe fiecare pereche de adjective se poate obţine un portret robot al grupului respectiv. Deoarece acest tip de scală are foarte multe din caracteristicile necesare unor prelucrări matematice riguroase, finalmente itemii valoroşi care au fost selectaţi pot fi asamblaţi într-o scală omogenă (de onestitate, de competenţă, de moralitate etc.).

Analogul vizual reprezintă un format foarte similar cu cel precedent, cu o singură diferenţă majoră: continuumul nu mai este segregat într-un număr discret de categorii de răspuns, ci subiectul marchează efectiv un semn pe linia continuă dintre cele două adjective, exact în locul unde crede că s-ar plasa opinia, credinţa sau experienţa sa. Ceea ce se înregistrează şi se prelucrează este distanţa măsurată faţă de unul dintre poli (se porneşte de regulă de la cel care înseamnă absenţa).

Nici urmă de frică _____x_____ Frică paralizantă

Deşi pare seducător, acest tip de scală are destul de multe dezavantaje. De multe ori precizia sa este o falsă precizie, marcarea punctului pe linie implicând şi o doză de hazard sau de circumstanţialitate. Un subiect va distinge cu greu între o frică de 6,7 cm de una de 6,5 cm. Mai mult, asignarea valorilor scalei poate fi una idiosincratică, adică distanţele marcate să aibă semnificaţii diferite de la un subiect la altul sau la acelaşi subiect la perioade de timp diferite.

Şi totuşi, aceasta poate fi un tip de măsurătoare foarte utilă atunci când cercetările se fac pe acelaşi subiect pentru a măsura diferenţele pe care tratamentul experimental le poate aduce în tratarea fobiilor, în creşterea nivelului de motivaţie sau în alte direcţii. Astfel, dacă scala Likert împarte un continuu într-un număr finit şi mic de categorii, fiind foarte maniabilă, dar mai puţin sensibilă, scala analogului vizual poate fi sensibilă la diferenţe mici: dacă răspunsurile la post-manipulare se depărtează semnificativ şi în aceeaşi direcţie în raport cu faza de pre-manipulare, se pot detecta diferenţele mici pe care le-a adus tratamentul experimental. Scala poate fi folosită şi ca una alcătuită dintr-un singur item, caz în care nu se mai poate distinge cu claritate între instabilitatea măsurătorii şi instabilitatea fenomenului măsurat. Prin urmare, cea mai bună strategie ar fi aceea de a dezvolta mai multe măsurătorii de tipul analogului vizual pentru acelaşi fenomen, de a le asambla într-o scală şi de a-i determina consistenţa internă (stabilitatea).

O ultimă problemă este cea legată de **cadrul temporal** la care pot face sau nu referire itemii unei scale. Preluând distincţia pe care a făcut-o Cattell în legătură cu unele dispoziţii de personalitate care au un caracter situaţional, în timp ce altele reprezintă structuri stabile de personalitate numite trăsături, Spielberger a dezvoltat scala STAI, în fapt o măsură a anxietăţii ca *stare* (conjuncturală) şi ca *trăsătură* de personalitate (stabilă). Se ştie de asemenea că multe componente ale personalităţii au mai multă sau mai puţină stabilitate pe axa timpului: fenomenologia afectivă de exemplu (emoţie, anxietate, depresie) este mai fluctuantă decât cea cognitivă (inteligenţă, memorie).

O scală trebuie să facă referinţe şi la **dimensiunea temporală**? Răspunsul este nuanţat: dacă constructul măsurat are o variabilitate temporală mare răspunsul este Da, dacă el este stabil în timp (cazul trăsăturilor de personalitate) răspunsul este Nu. Astfel, chestionarul de depresie Burns se aplică periodic pacienţilor depresivi pentru a monitoriza evoluţia depresiei lor în timp. Multe chestionare care urmăresc aspecte fluctuante menţionează în instructaj formule de tipul *în ultima lună, în ultima săptămână, în ultimele zile, în ultimul timp* etc. Unele chestionare includ itemi destinaţi detectării modificărilor de dispoziţie apărute recent (un eveniment stresant) sau pe perioade mai lungi de timp (reforma post-revoluţie, intrarea ţării în NATO etc.). Chiar formatul de răspuns al itemilor poate avea în vedere decuparea continuumului temporal în unităţi mai mici (*uneori, din când în când, des, în ultimul timp* etc.).

Înainte de a supune judecăţii experţilor forma preliminară a unei scale, constructorul trebuie să aibă în vedere dacă include sau nu şi **scale de validare**, care să diminueze sau să controleze tendinţa respondenţilor la dezirabilitate socială sau la fraudă. De exemplu Sandra Bem (BSRI, 1976), a inclus o scală de *dezirabilitate socială* ca tampon între itemii care măsoară masculinitatea şi cei care măsoară feminitatea. Eysenck a inclus în EPI (*Eysenck Personality Inventory*) o scală de minciună, devenită ulterior o scală de dezirabilitate socială în EPQ (*Eysenck Personality Questionnaire*).

În forma veche, dar şi în cea revizuită a MMPI (*Minnesota Multiphasic Personality Inventory*) Hathaway şi McKinley au propus mai multe scale de validare, care sunt primele ce vor fi evaluate de psiholog pentru a determina de cât credit se poate bucura profilul unui respondent. Problema validării se poate rezolva fie construind scale anume destinate acestui scop, fie preluând unele deja existente pentru a le insera în chestionar pe post de distractor şi de mijloc de control.

10.2.4. Pasul al patrulea: revizia setului iniţial de itemi de către experţi

Avizul experţilor este foarte necesar în legătură cu câteva probleme, sfatul lor scutindu-ne de cheltuieli mari de resurse pentru obţinerea de rezultate discutabile ca valoare. Expertiza acestora este necesară pentru a determina:

- faptul dacă definiţia şi operaţionalizarea fenomenului sau a constructului ce va fi măsurat sunt bune sau în acord cu ceea ce şi ei ştiu despre acestea;
- cât de relevanţi sunt itemii selecţionaţi pentru a fi incluşi în formatul iniţial al scalei, în raport cu constructul măsurat;
- claritatea, conciziunea şi alte elemente legate de modul de formulare al itemilor;
- evidenţierea fenomenelor sau a caracteristicilor importante asociate cu constructul ce va fi măsurat care nu s-au materializat în itemi corespunzători, setul iniţial trebuind deci augmentat cu noi itemi.

Prin revizuirea multitudinii de demersuri prin care vrem să surprindem constructul ce ne interesează şi prin examenul critic al itemilor destinaţi acestui scop experţii ne pot ajuta la maximizarea validităţii de construct şi de conţinut a scalei. Trebuie însă precizat că cel care ia deciziile finale este cel care construieşte efectiv scala, şi nu experţii. Este posibil ca faza pregătitoare pe care constructorul de scală a parcurs-o să-i fi dat un grad de cuprindere şi de înţelegere a problemei mai bun decât al fiecărui expert în parte, aşadar solicitarea lor de a fi

eliminaţi anumiţi itemi trebuie examinată cu atenţie pentru a vedea dacă ea are sau nu temei. În al doilea rând, cel mai adesea experţii vor tinde să elimine itemii redundanţi, eliminând astfel posibilitatea constructorului de scală de a alege şi de a păstra dintre cele două formulări aparent echivalente ale unor itemi, pe cea care se impune din analiza datelor rezultate din aplicarea instrumentului.

10.2.5. Pasul al cincilea: administrarea setului iniţial de itemi

Prima aplicare a unui instrument nou creat se face pe un lot suficient de larg pentru a da rezultate stabile referitoare la universul itemilor. Cât de mare şi de reprezentativ trebuie să fie aceste eşantion este în continuare obiect de dezbatere printre specialişti. Pentru o scală unidimensională, 250-300 de subiecţi sunt suficienţi pentru a desfăşura o analiză factorială, urmată de o analiză de itemi şi de scală, tehnici care vor ghida procesul de selecţie a itemilor ce evidenţiază cel mai bine constructul măsurat. Eşantionul iniţial se va mări în funcţie şi de numărul de subscale care vor fi extrase.

Dacă eşantionul este mic din punct de vedere numeric pot apărea erori prin şansă sau hazard, erori ce se vor repercuta asupra corelaţiilor dintre itemi şi scală, astfel că unii itemi, iniţial promiţători, să fie descoperiţi la retestare a fi de fapt unii slabi. Mai mult, loturile mici au o probabilitate mai mare de a elimina anumite categorii de indivizi cărora li se adresează scala respectivă. Pe de altă parte, dacă în eşantionul extras caracteristica măsurată este slab distribuită, valorile parametrilor obţinuţi (media şi abaterea standard în principal) sunt mai greu de extrapolat la populaţia generală. Gradul de generalizabilitate al acestora este în cazul de faţă unul redus.

De multe ori populaţiile la care avem acces (elevi de liceu, studenţi din campusuri etc.) sunt destul de diferite în privinţa caracteristicii măsurate faţă de populaţia ţintă a scalei. Unele diferenţe majore pot apărea din cauza conotaţiilor particulare pe care anumite cuvinte sau expresii le au într-o anumită categorie populaţională, generînd sensuri care nu au fost avute în vedere intenţionat de către constructorul scalei sau care nu se regăsesc ca atare şi în populaţia ţintă mai largă.

10.2.6. Pasul al şaselea: evaluarea itemilor din setul iniţial

Operaţiunile descrise la acest nivel sunt cele mai tehnice şi constituie inima întregului proces de construcţie al scalelor.

a. Examinarea iniţială a performanţei itemilor

Cea mai importantă calitate pe care o căutăm în această fază la un item este corelaţia sa ridicată cu variabila latentă pe care dorim să o evidenţiem prin constructul măsurat. Dar, cum această corelaţie nu poate fi determinată (dacă am şti valoarea ei nu ar mai avea rost să dezvoltăm o scală), noi putem face inferenţe în legătură cu aceasta plecând de la corelaţia dintre itemi şi de la corelaţia acestora cu scorul final la scală. Ca principiu, cu cât corelaţia dintre itemi este mai mare, cu atât mai mare este şi fidelitatea itemilor; cu cât itemii sunt mai fideli, cu atât mai mare este fidelitatea întregii scale.

Deci, prima calitate a unei scale este aceea de a fi una înalt intercorelată, lucru care se poate pune în evidenţă prin **matricea de intercorelaţii**. Un examen atent al acesteia relevă existenţa unor itemi care au corelaţii negative cu majoritatea celorlalţi itemi. Aceasta se poate întâmpla din mai multe cauze. De exemplu, itemii cu conţinut negativ care au fost scoraţi invers; itemii care, deşi scoraţi în acelaşi mod, măsoară constructe diferite (dominanţă vs. supunere, stabilitate emoţională vs. emotivitate). În principiu, aceştia ar putea fi reuniţi într-o scală supraordonată unidimensională, dar cu doi poli (bipolară). Pentru această categorie există mai multe soluţii. Prima ar fi aranjarea şi numerotarea descriptorilor invers în text, în funcţie de conţinutul itemilor, soluţie care nu este recomandabilă, pentru că nu este comodă pentru subiect. În timpul completării, regularitatea dispunerii itemilor în scală creează respondentului o anumită rutină prin care administrarea acesteia devine mai facilă. Inversarea distruge această regularitate, obligând respondentul la mai mult efort. A doua şi cea mai practică soluţie este aceea de a construi separat grile de corecţie pentru fiecare dintre dimensiunile măsurate de scală, prin care anumiţi itemi vor fi scoraţi direct iar alţii invers.

Pentru datele introduse în calculator în format de scală Likert, cea mai convenabilă soluţie este scorarea inversă a itemilor care ne interesează după formula: Nou = (x+1) - vechi. De exemplu, aplicând formula de mai sus pentru o scală cu 7 trepte, 7 devine 1, 6 devine 2 etc., iar 1 devine 7. În SPSS există şi opţiunea **Recode into the same variable**, prin care vechea variabilă este înlocuită cu cea rezultată după scorarea inversă. Avantajul este acela că se poate face transformarea simultană, în bloc, a tuturor variabilelor de acelaşi format (cu acelaşi număr de categorii de răspuns, definite în acelaşi fel).

Pentru că nu suntem întotdeauna siguri dacă rezultatul obţinut prin scorare inversă este mai bun decât cel precedent, este bine să lucrăm pe o copie a fişierului sursă, copie numită de exemplu „Itemi inversaţi”. Nu întotdeauna inversarea scorării, urmată de refacerea matricii de intercorelaţii, rezolvă problema unui item. Este posibil ca el să fie corelat pozitiv cu unii dintre itemi şi negativ cu alţii, deoarece el poate fi expresie a unui alt construct. În

acest caz soluţia potrivită este eliminarea itemului respectiv, pentru a nu periclita omogenitatea de ansamblu a scalei.

b. Corelaţia item-scală

În SPSS procedeul de lucru pentru acest tip de analiză este relativ simplu, deoarece se urmează următoarea secvenţă: **Analyse→Scale→Reliability→Alpha→Statistics**, de unde se activează **Item**, **Scale**; **Scale if item deleted** (scala, dacă itemul ar fi înlăturat). Dacă dorim să obţinem şi o matrice de intercorelaţii sau o matrice de covarianţe, în dreapta-sus avem panoul **Inter-Item**, de unde selectăm **Correlations** sau **Covariances** (vezi imaginea de mai jos). Finalmente dăm **Continue** pentru fereastra **Reliability Analysis: Statistics** şi apoi OK din fereastra **Reliability Analysis,** după ce vom fi introdus în panoul de lucru din dreapta-sus (**Items:**) toţi itemii care vor fi supuşi analizei.

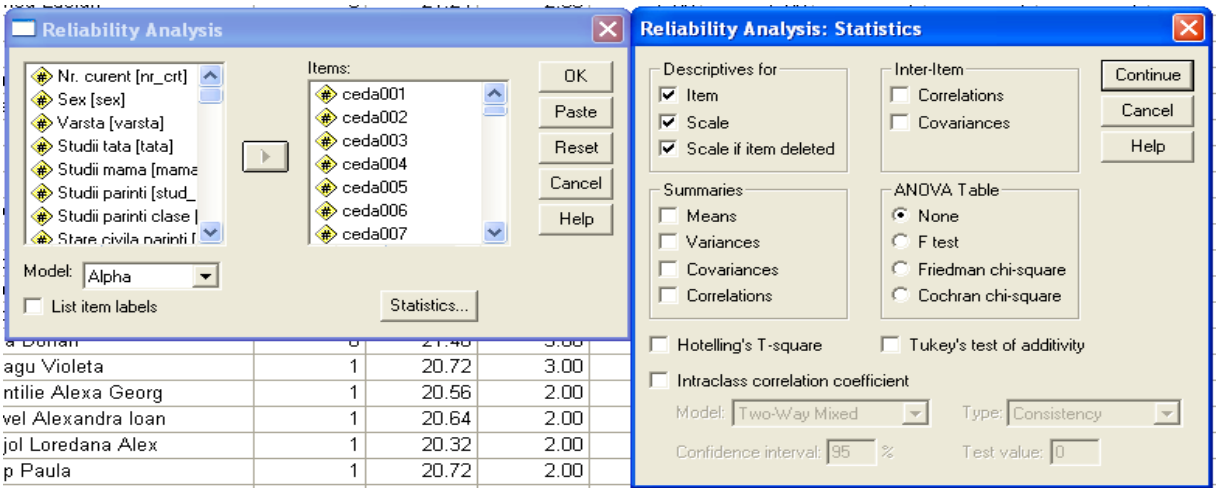

Figura 10.1. *Ferestrele de dialog pentru **Relyability Analysis** şi **Statstics**.*

Rezultatele arată la fel ca în fragmentele de output prezentate mai jos.

R E L I A B I L I T Y A N A L Y S I S - S C A L E (A L P H A)

		Mean	Std Dev	Cases
1.	CEDA001	.1500	.3601	60.0
2.	CEDA002	.1833	.3902	60.0
3.	CEDA003	.1167	.3237	60.0
4.	CEDA004	.3833	.4903	60.0
5.	CEDA005	.1667	.3758	60.0
6.	CEDA006	.5333	.5031	60.0
7.	CEDA007	.3500	.4810	60.0
8.	CEDA008	.3167	.4691	60.0
9.	CEDA009	.4667	.5031	60.0
10.	CEDA010	.1500	.3601	60.0

……

Item-total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
CEDA001	28.7500	133.4788	.2944	.8785
CEDA002	28.7167	133.4607	.2712	.8787
CEDA003	28.7833	135.0879	.1149	.8799
CEDA004	28.5167	133.4065	.2128	.8793
CEDA005	28.7333	133.1141	.3231	.8782
CEDA006	28.3667	136.6429	-.0713	.8825
CEDA007	28.5500	132.9975	.2550	.8788
CEDA008	28.5833	132.4845	.3105	.8782
CEDA009	28.4333	132.6226	.2746	.8786
CEDA010	28.7500	131.9534	.4804	.8770

…….....

Reliability Coefficients

N of Cases = 165 N of Items = 112

Alpha = .8800

……

Figura 10.2. *Outputurile rezultate din comenzile prezentate mai sus.*

Fiecare dintre fragmentele de output de mai sus trebuie examinate cu mare atenţie. Primul element investigat este **varianţa** itemilor, pentru că de aceasta depinde finalmente variabilitatea întregii scale. Dacă toţi indivizii investigaţi ar răspunde doar într-un singur fel la un item, varianţa (variabilitatea) sa ar fi zero, el nefiind de nici un folos în surprinderea constructului de măsurat. Cum unii subiecţi răspund într-un fel, iar alţii în altul, fiecare item are variabilităţi mai mari sau mai mici: cei mai dezirabili sunt cei cu variabilitate mare. În cazul de mai sus, itemii cu cea mai mare variabilitate sunt în ordine 6, 9, 4, 7 şi 8, care au o abatere standard în jur de 0,50.

Experienţa statistică ne arată că este mai bine să nu examinăm abaterea standard separat, făcând abstracţie de medie. În principiu, cu cât aceasta este mai apropiată de centrul intervalului de răspuns, cu atât este mai bine, pentru că ea va face posibilă mai multă variabilitate a itemului. În condiţiile în care la stânga au fost plasaţi descriptorii pentru dezacord şi la dreapta pentru acord, dacă media este asimetrică spre dreapta (negativ) înseamnă că majoritatea respondenţilor au ţinut să fie de acord cu conţinutul itemului respectiv, deci el a fost formulat „moale”; dacă asimetria este spre stânga (pozitivă) înseamnă că majoritatea şi-a exprimat dezacordul cu conţinutul itemului (formulare sa a fost cam „tare”).

În cazurile de asimetrie se ştie de asemenea faptul că puterea discriminativă a itemului este mai mare pe „coada” (creoda) curbei gaussiene şi mai mică pe ramura sa mai scurtă.

Cum o distribuţie simetrică (cu media pe centrul seriei de variaţie) are două cozi lungi, înseamnă că puterea discriminativă a unui item de acest fel este maximă, el putând distinge la fel de bine în ambele direcţii. În cazul nostru, deoarece numărul itemilor destinaţi depistării dezordinilor alimentare a fost foarte mare – 112 – am folosit scorarea dihotomică, deci valoarea medie optimă ar trebui să cadă la jumătatea intervalului 0 – 1, adică în jurul lui 0,50, ceea ce se întâmplă doar pentru itemii 6 şi 9.[27] Verificarea simultană a mediei şi abaterii standard este o condiţie prealabilă obligatorie pentru o tentativă de selecţie a itemilor care se bazează în principal pe corelaţie. Din această analiză reiese că, la prima vedere, itemii 7 şi 9 au simultan cele mai bune varianţe şi valori medii.

Decizia finală se ia însă după ce se analizează atent datele din panoul **Item-total Statistics** care are următoarea structură: primele două coloane dau valorile medii şi abaterile standard pentru valorile totale ale scalei, în situaţia în care itemul nu participă la scorul total. Pentru scalele cu număr mare de itemi corecţia adusă la parametrii scalei prin absenţa unei singure valori de scor este neglijabilă, dar ea devine cu atât mai importantă cu cât numărul itemilor scalei este mai mic, fiind importantă de exemplu pentru scala de 10 itemi. Coloana **Corrected Item-Total Correlation** este de maximă importanţă în luarea deciziei de selecţie deoarece ea indică gradul de asociere a fiecărui item cu scorul total, adică saturaţia sa în variabila latentă măsurată. În cazul nostru itemul 10 are o corelaţie excelentă cu scala (.48), în timp ce itemii 3, 4 şi mai ales 6 au corelaţii slabe (itemul 6 chiar negativă) cu scorul total.

Dacă scala noastră ar fi unidimensională (adică ar măsura doar un singur construct şi nu mai multe) o valoare a corelaţiei item-scală în jur de .30 ar fi suficientă pentru includerea itemului în scală, iar una de .40 aproape că ne forţează să facem acest lucru. Aici este cazul să spunem că nu ne vom concentra atenţia doar pe numerele de cod ale itemilor, fapt care ne împiedică să ştim care este conţinutul lor. Cu foaia de test în mână, după ce am identificat itemii cu formulări foarte asemănătoare (redundanţi), luăm în cunoştinţă de cauză decizia de a-i păstra pe cei cu calităţile psihometrice cele mai bune şi de a-i elimina pe cei mai slabi.

Ultima coloană are de asemenea un rol important în luarea deciziei relative la selecţia itemilor deoarece ne indică cât ar fi fost coeficientul de consistenţă internă alpha dacă itemul analizat ar fi fost eliminat. Prin compararea cu scorul de referinţă al acestui coeficient (.88 pentru toată scala iniţială, adică foarte mare) putem vedea dacă absenţa itemului ameliorează

[27] Trebuie menţionat că ipoteza răspândirii gaussiene a simptomelor tulburărilor alimentare în populaţia de bază este una care nu poate fi susţinută, deoarece o mulţime de itemi care determină anorexia şi bulimia sunt relativ rar întâlniţi la persoanele obişnuite. Aceştia itemi vor avea probabil o altă distribuţie pentru populaţiile speciale pe care scala va fi validată.

valoarea lui alpha (caz în care el trebuie eliminat) sau dimpotrivă, o scade (semn că este unul util). În cazul nostru itemul 6 este sigur unul condamnat, deoarece el dă o corelaţie uşor negativă cu scala şi nu este consistent cu aceasta (absenţa lui din scală îl măreşte pe alpha).

Şi totuşi nu suntem încă în măsură să luăm decizia finală pentru că nu am răspuns clar la întrebarea fundamentală dacă itemii testului măsoară sau bat în direcţia unei singure variabile (scală unidimensională) sau a mai multora (scală multidimensională, neomogenă). Supoziţia implicită pentru indicele de consistenţă internă alpha al lui Cronbach este aceea că toţi itemii măsoară o singură variabilă latentă, fapt ce nu este dat prin definiţie, ci trebuie atent verificat.

c. Analiza factorială exploratorie

În această fază a procesului de construcţie **analiza factorială exploratorie** devine obligatorie. Aşa cum se va observa când se parcurge până la capăt cursul de Statistică avansată, o analiză factorială este o tehnică pretenţioasă, pentru că are nevoie de un număr mare de subiecţi, de ordinul sutelor, pentru a furniza o soluţie stabilă şi bine adecvată datelor. Am solicitat din start un număr cel puţin mediu de subiecţi (250-300) pentru a efectua o analiză factorială corectă deoarece când este desfăşurată pe prea puţine cazuri, ea poate să dea valori instabile, ce compromit uneori iremediabil întregul proces de construcţie a scalei. Să reţinem deci că, în această fază de construcţie a scalei, determinarea prin analiză factorială a variabilei latente care stă în spatele unui set de itemi este un fapt crucial.

Mergând mai departe cu exerciţiul nostru prezentăm secvenţa de lucru pentru a efectua analiza factorială exploratorie: **Analyse→DataReduction→Factor**, care produce deschiderea următoarei fereastre de lucru, în care vom introduce (prima fereastră, panoul din dreapta-sus) toate datele de analizat, adică toţi itemii chestionarului nostru.

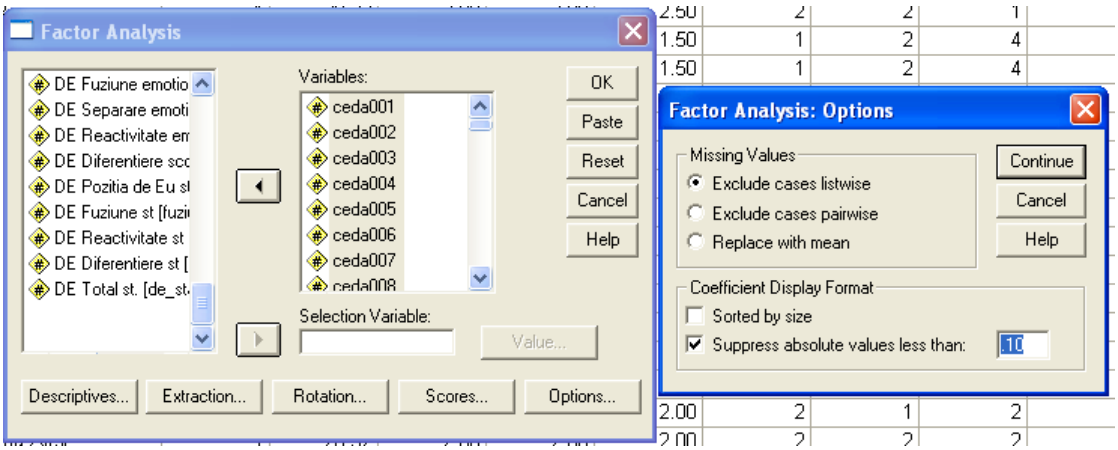

Figura 10.3. *Ferestrele de dialog pentru lansarea analizei factoriale exploratorii.*

Pentru un număr aşa de mare de itemi nu vom efectua nici o rotaţie (butonul din mijloc de la figura din stânga de mai sus), dar de la **Options** putem să selectăm **Supress absolute values less than:** (suprimarea valorilor mai mici decât:) pentru a pune valoarea de .30 în ferestruica respectivă, care are din setare valoarea .10, dar neactivată (vezi fereastra din dreapta-jos). În felul acesta valorile ce indică saturaţii mici şi nesemnificative în factorii descoperiţi nu vor mai fi afişaţi, fapt ce uşurează inspecţia vizuală, selecţia şi luarea deciziilor.

Trebuie precizat că metoda componentelor principale – care este opţiunea de bază a programului – este utilă mai ales pentru a vedea dacă există un factor general (cu o varianţă mai mare de 50%) ce defineşte variabilă latentă, sau dacă nu cumva există mai mulţi factori de magnitudine mai mică, ce acoperă întreaga varianţă a scorurilor scalei studiate. Vom putea vedea deci dacă avem de-a face cu o scală uni- sau multi-dimensională.

Atragem de asemenea atenţia asupra faptului că dacă o singură variabilă inclusă în panoul de lucru are varianţa zero, analiza factorială nu mai poate fi efectuată. Deoarece acesta este şi cazul chestionarului nostru, identificăm variabila sau variabilele cu varianţă zero prin **Frequencies** ori **Descriptives.** În cazul nostru itemul 66 trebuie eliminat din analiză (şi din scală deoarece, având varianţa zero, el nu are nici o utilitate), după care reluăm analiza factorială. Aceasta va produce mai multe output-uri, din care două sunt esenţiale.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	14.827	13.357	13.357	14.827	13.357	13.357
2	9.110	8.208	21.565	9.110	8.208	21.565
3	5.501	4.956	26.521	5.501	4.956	26.521
4	5.035	4.536	31.057	5.035	4.536	31.057
5	5.010	4.514	35.571	5.010	4.514	35.571
6	4.298	3.872	39.443	4.298	3.872	39.443
7	4.051	3.650	43.093	4.051	3.650	43.093
8	3.830	3.450	46.543	3.830	3.450	46.543
9	3.438	3.097	49.640	3.438	3.097	49.640
10	3.242	2.921	52.561	3.242	2.921	52.561
11	3.014	2.715	55.276	3.014	2.715	55.276
12	2.896	2.609	57.885	2.896	2.609	57.885
13	2.746	2.474	60.359	2.746	2.474	60.359
14	2.630	2.369	62.728	2.630	2.369	62.728
15	2.426	2.186	64.914	2.426	2.186	64.914
16	2.314	2.084	66.998	2.314	2.084	66.998

Figura 10.4. *Unul dintre outputurile analizei factoriale indicând factorii selecţionaţi şi valorile lor* ***eigen.***

Outputul de mai sus indică faptul că itemii iniţiali nu produc un singur factor general, ci câteva zeci (fapt curent în construcţia de scale cu număr mare de itemi), din care primul acoperă doar 13,36% din varianţă (deci este foarte departe de a fi un factor general), al doilea 8,21%, al treilea 4,96% etc. Chiar şi al şaisprezecelea factor produce o valoare **eigen** peste 2 (2,08%), ceea ce poate părea un adevărat coşmar statistic. Lucrurile nu stau însă chiar aşa. Dacă analizăm cu atenţie outputul al doilea din figura 1.4, observăm că acesta devine o adevărată grilă de selecţie pentru factorii mai importanţi care vor fi reţinuţi. Astfel, itemii vor fi alocaţi la acei factori pentru care ei au cele mai mari corelaţii (saturaţii în factorul respectiv). Atunci când se poate, itemii cu saturaţii negative se alocă la factorul cu care corelează pozitiv mai semnificativ, pentru a nu schimba maniera de scorare; la valori foarte apropiate ale saturaţiilor, un item se alocă scalei mai scurte pentru a o fortifica, deoarece ştim că puterea unei scale depinde şi de numărul itemilor care o compun.

Observăm că, după ce încep să fie alocaţi la primii trei factori, mai rămân foarte puţini itemi sau deloc pentru factorii următori, şi această deoarece mulţi itemi îşi împart varianţa între mai mulţi factori (nu sunt foarte specifici). Alocând itemii factorului pentru care ei au cea mai mare saturaţie, celorlalţi factori le rămân tot mai puţini itemi, aceasta deoarece primii factori dau corelaţii mai mari cu itemii, având deci prioritate în selecţia acestora. De fapt afişarea în tabel a factorilor de la stânga la dreapta se face în ordinea magnitudii lor, dată de mărimea corelaţiilor cu itemii, dar şi de numărul acestor corelaţii. Itemii care nu se repartizează în primele 3-4 categorii vor putea defini şi ei nişte factori de magnitudine mai redusă, dar reunindu-i în scale şi studiind alpha pentru acestea vom observa că şi consistenţa internă este din ce în ce mai mică pentru factorii extraşi ultimii.

	1	2	3
CEDA001			
CEDA002			
CEDA003		.343	.331
CEDA004			
CEDA005	.340		
CEDA006			
CEDA007	.324		
CEDA008	.303		
CEDA009		.343	
CEDA010	.530		
CEDA011			
CEDA012			
CEDA013	.539		
CEDA014			
CEDA015	-.458		.504
CEDA016	.442		
CEDA017	.509		
CEDA018			
CEDA019	.392	.417	-.364
CEDA020	.613		-.402
CEDA021	.327		
CEDA022	.429		-.444

Figura 10.5. *Output al analizei factoriale indicând factorii selecţionaţi şi valorile lor* ***eigen****.*

Cum o scală trebuie să aibă o consistenţă internă de peste .70 (aceasta fiind foarte bună peste .80 şi excelentă peste .90) scalele mici neatingând aceste valori trebuie în principiu abandonate, pentru că ele sunt expresia unor construncte ce au o valoare îndoielnică. În cazul de mai sus, itemii 5, 7, 8, 10, 13, 16, 17, 20, 21, 22 etc. vor fi alocaţi primei scale; itemii 3, 9, 19 etc. celei de a doua, iar itemii 15, 24, 43 etc. celei de a treia. O atenţie specială impune itemul 3, care poate fi distribuit şi la factorul 2 şi la 3 (decizia o luăm ulterior, când evaluăm care dintre cele două subscale are mai mare nevoie de el) sau la itemul 22, care corelează semnificativ şi cu factorul 1 (pozitiv) şi cu factorul 3 (negativ). În aceste situaţii se observă cât de importantă este mărimea lotului pentru luarea unei decizii corecte, bazată pe analiza factorială.

Hotărârea de a aloca itemii într-o scală sau alta poate fi irevocabilă, chiar dacă datele ulterioare nu confirmă deciziile noastre de acum. Prin mărirea semnificativă a numărului de subiecţi, soluţia factorială poate să aducă o mulţime de diferenţe, ceea ce relansează practic întreg procesul de reconstrucţie a scalei. Din această cauză vom căuta să luăm decizii importante şi definitive doar dacă eşantionul pe care s-a efectuat analiza factorială este unul sufiucient de mare şi de reprezentativ.

Finalmente, scala desemnând factorul 1 arată ca în figura de mai jos.

R E L I A B I L I T Y A N A L Y S I S - S C A L E (A L P H A)

		Mean	Std Dev	Cases
1.	CEDA005	.2000	.4031	65.0
2.	CEDA007	.3231	.4713	65.0
3.	CEDA008	.3231	.4713	65.0
4.	CEDA010	.1538	.3636	65.0
5.	CEDA013	.4154	.4966	65.0
6.	CEDA016	.5385	.5024	65.0
7.	CEDA017	.4154	.4966	65.0
8.	CEDA020	.1231	.3311	65.0
9.	CEDA021	.4462	.5010	65.0
10.	CEDA022	.3385	.4769	65.0

.....

Statistics for	Mean	Variance	Std Dev	N of Variables
SCALE	11.9846	86.4841	9.2997	50

Item-total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
CEDA005	11.7846	84.0154	.3121	.9203
CEDA007	11.6615	83.3212	.3418	.9202
CEDA008	11.6615	83.4462	.3270	.9204
CEDA010	11.8308	83.3928	.4456	.9192

CEDA013	11.5692	81.3115	.5500	.9180
CEDA016	11.4462	82.5635	.4018	.9197
CEDA017	11.5692	81.7178	.5034	.9185
CEDA020	11.8615	83.4337	.4862	.9190
CEDA021	11.5385	83.8149	.3636	.9202
CEDA022	11.6462	83.1385	.3586	.9201

.....

Reliability Coefficients
N of Cases = 165.0 N of Items = 50

Figura 10.6. *Unul dintre outputurile analizei factoriale indicând factorii selecţionaţi şi valorile lor* ***eigen****.*

d. Scurt raport despre scală

Se pare că scala 1 obţinută de noi este una foarte bună deoarece la ea contribuie 50 de itemi ce-i dau o consistenţă internă extrem de mare (alpha = 0,92). Fiecare item dă corelaţii de peste 0,30 cu scala, dar şi mai mari. Scorul mediu la această scală este de 11,98, cu o abatere standard de 9,30. Totuşi valorile medii ale itemilor nu sunt centrate majoritar pe valoarea mijlocie de scor (0,50), ci sunt mai mici, fapt ce indică o asimetrie rezultată din cauza caracterului mai curând clinic al testului. Aceşti itemi discriminează bine printre subiecţii care au valori mai mari ale scorurilor, dar mai slab pentru cei cu valori submedii. Scala urmează a fi verificată şi validată contra criteriul extern pe populaţii clinice care prezintă realmente dezordinile alimentare diagnosticate.

Studiind itemii care îl compun, primul factor a fost numit *Stimă de sine negativă* legată de aspectul fizic (50 de itemi), al doilea *Propensiune anorexică* (12 itemi) şi al treilea *Propensiune bulimică* (12 itemi). Deoarece cei trei factori sunt mediu-slab corelaţi între ei, scorurile lor separate pot fi agregate într-un scor global, adică într-un Index al tulburărilor alimentare, definit de preocuparea pentru aspectul fizic şi pentru pierderea controlului asupra ingestiei de alimente.

Numărul relativ mic de subiecţi pe care s-a desfăşurat analiza factorială (N = 165) ridică totuşi semne de întrebare în legătură cu stabilitatea soluţiilor găsite, care trebuie verificate şi definitivate pe populaţii mai largi.

10.3. Criterii de selecţie a itemilor pentru optimizarea lungimii scalei

Până în acest stadiu de dezvoltare al scalelor de tulburări alimentare cercetătorul are un număr bun de itemi (72), distribuiţi pe trei subscale în raport cu care aceştia au o fidelitate

destul de ridicată. Şi totuşi, se observă diferenţe relativ însemnate între acestea: prima are un alpha de .92, a doua de doar de .70, iar a treia de .58. Alpha pentru toată scala este de .94. Se pare deci că scalele 2 şi 3 nu sunt suficient de fidele, aceasta probabil şi din cauza scurtimii lor. Se admite că scalele scurte sunt bune, pentru că se completează uşor şi nu pun prea mare presiune pe subiecţi. Prin opoziţie, scalele mai lungi creează mai mult stres la completare, dar au caracteristici psihometrice mai bune (consistenţă internă, alte forme de fidelitate şi de validite, putere de discriminare mai mari). Este deci de dorit găsirea unui echilibru între scurtimea scalei şi calităţile sale psihometrice.

Atunci când fidelitatea unei scale este prea mică, scurtimea ei nu este o calitate în sine. Dacă cercetătorul nu poate asigura nici un înţeles unei scale scurte, nu s-a câştigat nimic prin scurtimea ei. În cazul de faţă există o satisfacţie legată de calitatea primei scale. Pentru cea de a doua şi de a treia ar putea reîncepe travaliul de a genera itemi suplimentari, care să surprindă mai specific bulimia şi anorexia şi care vor trebui aplicaţi unor populaţii clinice identificate a avea aceste afecţiuni.

În rezumat, pentru a clarifica criteriile după care păstrăm unii itemi în scală şi-i eliminăm pe alţii, furnizăm în sinteză cele mai importante repere orientative:

- sunt preferabili itemii care au valori medii apropiate de valoarea centrală, au varianţă mare, se corelează bine cu ceilalţi itemi şi – mai ales – corelează la un nivel ridicat (peste .30, sau chiar peste .40) cu scorul total la scală;
- vor fi primii eliminaţi itemii care dau varianţa zero, corelaţii slabe sau negative cu scala (chiar după scorarea inversă);
- itemii care corelează slab cu scala vor fi de asemenea printre primii eliminaţi;
- vor fi eliminaţi şi itemii a căror prezenţă în scală nu numai că nu ameliorează consistenţa sa internă, ci chiar o coboară;
- itemii care corelează slab cu dimensiunea/dimensiunile latente măsurate, dar corelează bine cu alţi factori secundari, vor fi şi ei eliminaţi din scală, dar vor fi analizaţi pentru a dezvolta eventual noi scale, dacă ei par să sugereze existenţa altor constructe valide;
- dintre itemii foarte similari ca şi formulare şi conţinut se păstrează doar aceia care dau cea mai bună corelaţie cu scorul total la scală.

Nu putem omite o întrebare legitimă: cât de mic trebuie să fie pragul corelaţiei item-scală pentru a reţine un item în scală? Răspunsul depinde şi de intenţiile autorului: doreşte el o scală scurtă şi relativ puternică, sau o scală mai lungă, dar care dă mai multe nuanţe, fiind mai stabilă şi mai fidelă? În primul caz pragul de selecţie este mai ridicat (în principiu peste r = .40), în al doilea ceva mai scăzut (dar rareori sub r = .30).

Efectul eliminării sau includerii unor itemi în scală trebuie şi el bine ştiut. Dacă cunoaştem corelaţia medie a scalei rezultate putem aplica o formulă care ne ajută să anticipăm care va fi efectul adăugării sau eliminării unui item. Această formulă este:

$$alpha = \frac{N \cdot r}{1 + (N - 1) \cdot r}$$

în care N este numărul de itemi şi r corelaţia medie inter-itemi. Putem astfel determina ce se întâmplă dacă la o scală cu corelaţia iter-itemi de .30 adăugăm sau scoatem un item: alpha = [10·.30]/[1+9·.30] = .81 pentru 10 itemi; alpha = [11·.30]/[1+10·.30] = .83 pentru 11 itemi; alpha = [9·.30]/[1+8·.30] = .79 pentru 9 itemi. Vedem că diferenţa de doar doi itemi ridică consistenţa internă de la .79 la. 83, ceea ce este important. În general, dacă corelaţia inter-itemi a unui item este egală sau cu puţin sub corelaţia inter-itemi medie, păstrarea itemului în scală are un efect de creştere a lui alpha; când această corelaţie este mult sub medie, ea poate să ducă la descreşterea lui alpha şi deci itemul rău trebuie eliminat.

O soluţie interesantă pentru a verifica stabilitatea caracteristicilor unei scale este propusă de DeVellis (2003, pp. 99-100), ea constând în splitarea lotului. Când lotul este unul suficient de mare el poate fi împărţit în jumătăţi. Când lotul nu este suficient de mare, împărţirea se poate face în părţi inegale, partea cea mai mare fiind rezervată pentru determinările principale, pe când cea de a doua pentru verificarea constanţei rezultatelor găsite. Deci prima parte din lot foloseşte la evaluarea şi selecţia itemilor, la construcţia scalelor şi la determinarea lui alpha pentru fiecare subscală; cea de a doua parte este utilă pentru replicarea datelor şi pentru a verifica dacă ele se regăsesc pe un lot foarte similar.

Avantajele acestei metode – care obligă din start la abordarea unui număr mare de subiecţi, de peste 500-600 de persoane – sunt însemnate. În primul rând împărţirea se poate face aleatoriu, ceea ce asigură echivalenţa grupurilor, randomizarea generând două eşantioane echivalente. Dacă acestea ar fi abordate în două momente temporale diferite există probabilitatea ca factorul timp să aducă modificări (fie şi discrete) la cel de al doilea grup, sau ca el să difere prin alte caracteristici de primul grup. Apoi, un nou grup presupune expunerea subiecţilor la alt personal de cercetare, în altă perioadă a anului, cu instrumente care pot diferi uşor de primele (forma iniţială şi forma definitivă a scalei), ceea ce poate introduce alte

distorsiuni nedorite. Dar cel mai însemnat avantaj al splitării grupului original este acela că la ambele subgrupuri vom avea forma originală a instrumentului, ceea ce permite revenirea la decizia eliminării unor itemi dacă grupul de control indică necesitatea schimbării acestei decizii. Dacă pentru primul grup, cel pe care s-a făcut selecţia itemilor, există posibilitatea ca factori de şansă să fie confundaţi cu covariaţia itemilor, la al doilea – unde itemii sunt deja selectaţi – o asemenea posibilitate nu mai există. Finalmente, prin reunirea celor două grupuri în unul singur se obţin caracteristici psihometrice mai sigure şi mai aproape de realitate decât pentru fiecare grup luat separat.

10.4. Alte analize utile pentru determinarea fidelităţii scalei

Ideea că prin coeficientul alpha al lui Cronbach se rezolvă problema fidelităţii unui test/scale este una eronată deoarece în esenţă acesta indică doar măsura în care itemii formează un set omogen, bine sau satisfăcător corelat cu scorul total la test. Alpha nu ne spune totuşi dacă scala nou creată surprinde sau nu cu adevărat estenţa constructului căutat, pentru aceasta trebuind desfăşurate nenumărate studii de validare, pentru oricare nouă extensie legată de utilizarea scalei. Chiar fidelitatea presupune determinarea prin test-retest a stabilităţii în timp a rezultatelor obţinute, corelaţia dintre o formă a testului şi forma sa paralelă etc. Cum fidelitatea este cea mai importantă precondiţie a validităţii, trebuie spus că programul SPSS oferă alături de alpha şi posibilitatea determinării fidelităţii prin metoda jumătăţirii (**split-half**).

Demersurile necesare acestui tip de analiză sunt foarte asemănăroare cu cele presupuse de determinarea lui alpha (**Analyse→Scale→Reliability→Split-half→Statistics**), doar că de la **Statistics** acum se dezactivează **Descriptives for** (**Item**, **Scale**, **Scale if item deleted**), deja determinate anterior.

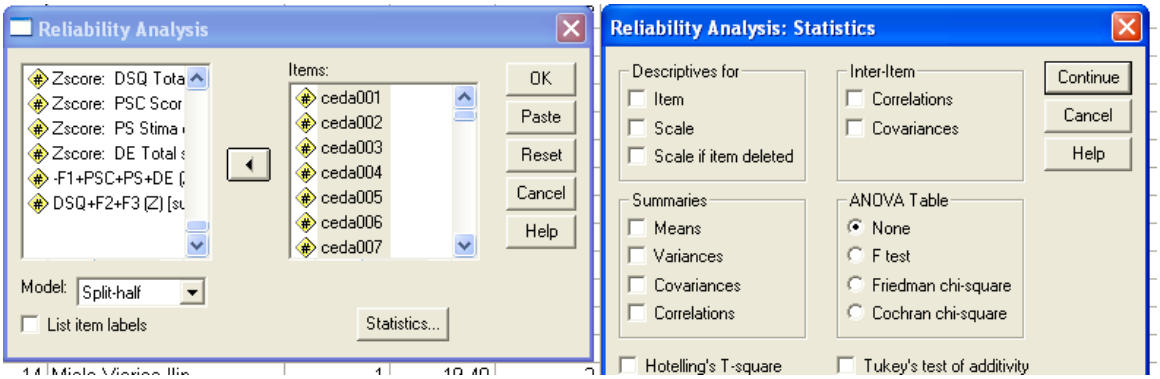

Figura 10.7. *Ferestrele de dialog pentru* ***Split-half*** *şi* ***Statistics***.

Output-ul acestei serii de comenzi este un indicator al consistenţei interne care ia în calcul o jumătate a testului comparată cu cealaltă, prin două metode: cea a lui Spearman-Brown şi cea a lui Guttman. Echivalenţa jumătăţilor se stabileşte şi prin compararea coeficientului alpha pentru fiecare jumătate în parte. Cele două jumătăţi comparate sunt luate de computer ca atare, dar dacă vrem să determinăm split-half pentru numerele pare şi cele impare, introducem în panoul de lucru întâi itemii pari, apoi pe cei pari, dăm OK şi vom obţine analiza dorită.

R E L I A B I L I T Y A N A L Y S I S - S C A L E (S P L I T)

Reliability Coefficients	
N of Cases = 165	N of Items = 112
Correlation between forms = .80	Equal-length Spearman-Brown = .89
Guttman Split-half = .89	Unequal-length Spearman-Brown = .89
56 Items in part 1	56 Items in part 2
Alpha for part 1 = .85	Alpha for part 2 = .87

Figura 10.8. *Rezultatele outputului pentru* ***Split-half****.*

Corelaţiile obţinute prin splitare sunt identice prin cele două metode (.89) iar valorile alpha pentru prima parte (.85) şi pentru cea de a doua (.87) sunt extrem de asemănătoare, de unde deducem că există o foarte bună consistenţă internă pentru fiecare jumătate a testului.[28]

Procesul de construcţiei a unei scale nu se opreşte aici. Destinul acesteia depinde doar parţial de calităţile psihometrice identificarte preliminar şi mai mult de utilitatea ei reală, de posibilitatea implicării ei într-o multitudine de aplicaţii practice sau de cercetări. Cel mai adesea validitatea este cea care impune în timp o scală şi aceasta poate fi probată într-o multitudine de maniere. Prezentăm mai jos o matrice de intercorelaţii dintre cei trei factori rezultaţi pentru *Scala de tulburări alimentare* şi cele trei categorii de mecanisme de apărare decelate de DSQ (*The Defense Style Questionnaire)* creat de Andrews, Singh şi Bond.

	FACTOR1	FACTOR2	FACTOR3	DSQ Mec. mature	DSQ Mec. nevrotice
FACTOR1	-				
FACTOR2	.38**	-			
FACTOR3	.32**	.28**	-		
DSQ Mec. mature	-.09	.16	.01	-	
DSQ Mec. nevrotice	.23*	.10	.18	.19*	-
DSQ Mec. imature	.51**	.31**	.29**	.21*	.53**

Notă: ** Corelaţia este semnificativă la pragul de 0,01 (bidirecţional).
* Corelaţia este semnificativă la pragul de 0,05 (bidirecţional.

Figura 10.9. *Matrice de intercorelaţii dintre cei trei factori identificaţi şi DSQ.*

[28] **Notă**: Aceste determinări au fost făcute pe tot setul de itemi doar ca exerciţiu, înainte de selecţia şi sortarea pe scale a itemilor, ele trebuind reluate pentru subscalele obţinute în finalul procesului de construcţie, ca şi pentru scala globală care le include pe toate trei. Dublarea numărului de subiecţi pentru a obţine o soluţie factorială stabilă relansează practic întreg procesul de reconstrucţie al scalelor.

Din tabelul de mai sus rezultă că factorul 1 se validează la un nivel foarte ridicat prin corelare cu mecanismele de apărare imature (r = .53) şi la un nivel mai scăzut cu cele nevrotice (r = .23). Aceasta atrage atenţia asupra faptului că factorul 1 are un conţinut negativ şi că ar trebui să se numească *Stimă de sine* ***negativă*** legată de aspectul fizic. Factorii 2 şi 3 se validează şi ei concurent cu mecanismele de apărare imature, dar la niveluri mai scăzute (r = .31, respectiv r = .29). Evident că această primă confirmare are nevoie de o mulţime de alte studii, dar aceasta presupune ca instrumentul să fie publicat şi lansat pentru a-şi îndeplini rolurile pentru care el a fost proiectat.

10.5. Proiect aplicativ:
construirea unui chestionar pentru o trăsătură de personalitate

Deoarece o treime din nota finală depinde de acest proiect, cealaltă depinzând de activitatea de seminar şi ultima de nota de la examenul final, indicăm mai jos paşii care trebuie urmaţi pentru realizarea sa:

- veţi citi cu atenţie lista de la sfârşitul paragrafului de faţă şi veţi alege o trăsătură de personalitate, de preferinţă una care vă interesează personal;
- veţi efectua un studiu bibliografic detaliat pentru a putea identifica şi defini detaliat trăsătura/dimensiunea de personalitate aleasă (aceasta este faza de documentare);
- folosind metoda intuitiv-raţională veţi operaţionaliza această trăsătură/dimensiune aleasă sub forma unui număr de itemi cuprins între 20 şi 40. În construcţia itemilor veţi ţine cont de toate precizările din cursul de faţă legate de caracteristicile de formă, lungime, conţinut şi cele semantice;
- chestionarul astfel obţinut va fi aplicat într-un studiu-pilot pe un număr de minimum 30 de subiecţi. Dacă optaţi pentru strategia grupurilor-criteriu (metoda empirică), acestea vor avea minimum 20 de subiecţi fiecare;
- utilizând cunoştinţele de la cursul *Bazele teoretice ale psihodiagnosticului* veţi efectua analiza itemilor (grad de discriminare, curbă caracteristică sau semnificaţia statistică a diferenţei dintre grupurile de contrast) şi veţi reţine în forma intermediară redusă a chestionarului itemii cu caracteristicile psihometrice cele mai bune. Pentru această formă a chestionarului veţi urma procedura descrisă în curs şi veţi face analiza

consistenţei interne a scalei (din SPSS →**Analyse→Scale→Alpha Cronbach** şi apoi **Split-half**);

- veţi redacta forma finală a chestionarului;
- opţional, valoarea proiectului va fi considerată mai bună şi va primi punctaj superior dacă veţi corela produsul obţinut de voi cu un instrument despre care se ştie că măsoară acelaşi construct, pentru a proba validitatea concurentă a chestionarului vostru.

Notă: datorită complexităţii cerinţei formulate vă puteţi asocia câte doi studenţi pentru un singur proiect, caz în care nota la acesta va fi identică pentru ambii autori, deşi ei îşi pot împărţi după dorinţă atribuţiile între faza de documentare şi cea de aplicare/prelucrare a datelor. În acest caz însă ultimul aspect (validarea concurentă a chestionarului nou creat) devine obligatorie.

Produsul finit (chestionarul împreună cu o lucrare de prezentare a fazelor proiectului, de 3-4 pagini în TNR 12, pe hartie A4) va fi însoţit de o dischetă sau CD cu datele de pe calculator şi prelucrările aferente. Data predării proiectului: ultima oră de curs din semestru.

Prezentăm mai jos câteva concepte care descriu trăsături de personalitate sau alte domenii sau teme de cercetare ce pot fi abordate prin secvenţa metodologică prezentată anterior:

Adaptare (şcolară, universitară, organizaţională), agresivitate, asertivitate, atitudine faţă de (alcoolism, consum de droguri, risc, sexualitate, carieră etc.), altruism/egoism, alimentare (bulimie, anorexie, bulimi-anorexie), ambiţie, androginitate, angoasă/anxietate/fobie socială, aptitudini (didactică, socială, de lider, empatică etc.), ataşament (faţă de valorile organizaţiei, faţă de tradiţie), aviditate (ca în chestionarul Gaston-Berger din tipologia franco-olandeză), bătrâneţe, bioenergie/bioenergoterapie (atitudine faţă de), bulimie, cafea, carenţă afectivă, carismă, chirologie, cinema, cleptomanie, comandă, competiţie, complexe (de inferioritate sau de superioritate), comportament matern/patern, comportament proactiv, comunicativitate, conflict/ conflictualitate (toleranţă la), conformism/nonconformism, contagiune (afectivă şi mentală), control (locus of), cooperare, copil/copilărie (atitudine faţă), creativitate, culpabilitate, decizie (uşurinţă/dificultate), diferenţiere (a Eului, cognitivă), delincvenţă (tendinţă la), dependenţă (alcoolică, de droguri, emoţional-afectivă), depresie, dezacord/acord conjugal, didactogenie, dipsomanie, disforie, dispoziţie spre (tutun, alcool, droguri, abuz

sexual), drog, durere (căutarea plăcerii/evitarea durerii - Cloninger), educaţie/ educabilitate, efort (rezistenţă la), egocentrism/egoism, emotivitate, empatie, erotism/senzualitate, eşec (toleranţă/intoleranţă, conduită de eşec), Eu ideal, expectaţie (nivel de), extraversiune/introversiune, fabulaţie (gândire magică), familie (fericire/nefericire conjugală sau matrimonială), feminitate, fidelitate (relaţională şi în căsnicie), foamea de stimulare, fobii (sensibilitate la), frică, frigiditate, frustrare (toleranţă la), furie, gândire (magică, negativă, depreciativă, autodevalorizatoare, creativă etc.), gelozie, gregarism, grup (atitudine faţă de), gust, hiperkinezie, homosexualitate/androginitate/sex-rol, idei (de persecuţie, prevalente, dominante) şi scheme, identificare, imaginaţie (reproductivă şi creatoare), imitaţie, inadaptare la (mediu, şcoală, serviciu, căsnicie), impulsivitate, inhibiţie, instabilitate (emoţional-afectivă, caracterială), interese, introversie, intuiţie, isteroidie, iubire, izolare, joc (disponibilitate pentru), încredere (în sine şi în alţii), labilitate, lateralitate corticală, lene socială, lider (aptitudine de), magie (gândire magică), manie, masochism, mecanisme de apărare ale Eului, melancolie (dispoziţie spre), minciună (scală de), mitomanie, modă (atitudine faţă de), moarte (atitudine faţă de), motivaţie (şcolară, profesională, de muncă, creatoare etc.), narcisism, nebunie (teama de), negativism/încăpăţânare, nervozitate, neutralitate (neimplicare), nevroză (nevrozism, tendinţe nevrotice), nevroză de abandon, nevroză şcolară, obezitate, oboseală (existenţială), obsesii, opinii (de orice fel), optimism, orgoliu, panică (atacuri de), paranoia (tendinţe spre), parapsihologie (atitudine faţă de), pasiune, pasiv-dependentă (personalitate), pedeapsă (atitudine faţă de), pensionare, perfecţionism/introversie, personalitate (isterică, astenică, fobică etc.), persuasiune/sugestibilitate, perversiune (atitudine faţă de), placebo (susceptibilitate la), pragmatism (la valori/atitudini fundamentale), prejudecăţi, prestigiu, prezenţă socială, privaţiune (toleranţă la), pubertate/adolescenţă, răsunetul reprezentărilor, regresie, relaţii umane, relaxare, remuşcare, respingere, responsabilitate, reuşită/succes (şcolar, profesional etc.), reverie, revoltă, rezistenţă, rezonanţă, ritualuri fobice, roluri, sadism, sănătate mintală, schizoidie, secundaritate, sensibilitate, sentiment, separare (nevroză de), sete (afectivă, de cunoaştere), sfat conjugal, simulare (comportament), suprasolicitare/burnout/brownout, sinucidere, sofrologie, stîngăcie, stres (toleranţă la, mecanisme de coping cu), subiectivitate, succes, sugestibilitate, surmenaj, şoc, talent, tandreţe, temperament, tensiune, ticuri, timp (atitudine faţă de, gestiune a), toleranţă (la frustrare, socială), trac, tutun, uitare, umor, valenţe, valori, vârstă (atitudine faţă de), veleitarism, virilitate/masculinitate, vis, vocaţie, voinţă, vorbire, zvon.

CAPITOLUL 11

RAPORTUL PSIHOLOGIC

11.1. Obiective

După parcurgerea acestei unităţi, studenţii vor fi capabili:

- să cunoască elementele de conţinut ale celor opt capitole ce intră în structura raportului psihologic;
- să facă diferenţele care se impun între *fişa psihologică* (care sintetizează cifric datele clientului) şi *raportul psihologic*, care insistă pe interpretarea narativă a acestora;
- pentru fiecare dintre cele opt capitole din structura raportului psihologic să poată opera definirea, indica funcţiile, selecta elementele necesare realizării corecte şi evitării erorilor celor mai probabile;
- să concretizeze toate informaţiile dobândite în blancul unei foi de raport-tip, printr-un exerciţiu de portofoliu.

11.2. Lecturi recomandate pentru acest curs

1. Aiken, L.R. (1997). *Psychological Testing and Assessment*. 9th ed. Boston, London, Toronto, Sydney, Tokyo, Singapore: Allyn and Bacon.
2. Albu, M. (1998). *Construirea şi utilizarea testelor psihologice*. Cluj-Napoca: Editura „Clusium".
3. Anastasi, A. (1974). *Psychological Testing*. New York: Macmillan Publishing Co. Inc.
4. Clinciu, A.I. (2005). *Psihodiagnostic*. Braşov: Editura Universităţii Transilvania.
5. Gregory, R.J. (1996). *Psychological Testing. History, Principles and Applications*, second edition, Boston, London, Toronto etc: Allyn and Bacon.
6. Stan, A. (2002). *Testul psihologic. Evoluţie, construcţie, aplicaţii*. Iaşi: Editura Polirom.
7. Zörgo, B, (1976). „Examinările psihologice". În *Îndrumător psihodiagnostic*. Cluj-Napoca: Reprografia Universităţii Babeş-Bolyai, pp. 1-16.

11.3. Materiale şi suporturi necesare

Hârtie şi creioane; fişă (blanc) de raport psihologic; câteva fişe psihologice şi un raport psihologic real, rezultate din practica psihologică, de model.

11.4. Idei de bază pentru întocmirea unui raport psihologic

1. Raportul psihologic este expresia condensată a activităţii de testare / evaluare psihologică a unei persoane (client, consumator de servicii etc.) de către un psiholog calificat care răspunde cu instrumentele ştiinţei sale scopului pentru care a fost desfăşurată investigaţia, în termeni profesionişti, dar şi accesibili celui care a comandat-o.

2. Raportul psihologic nu se confundă cu Fişa psihologică, deoarece prima ramâne la client, a doua la psiholog; prima "traduce" în termeni inteligibili informaţia prezentă în fişă, la care se adaugă obligatoriu capitole de interpretare, sumarizare (concluzii) şi recomandări.

3. Raportul psihologic nu exclude partea tehnică de prezentare a datelor (notele brute şi notele standard ale testelor utilizate), dar importantă este "traducerea" limbajului cifric într-o descriere narativă, într-un limbaj accesibil clientului.

4. Ca structură, raportul psihologic (sau de examinare psihologică) cuprinde datele factuale - de identificare, scopul sau problema pentru care a fost solicitată testarea, testele administrate şi rezultatele obţinute, exprimate sub formă de note brute (eventual) şi note standard (obligatoriu), o anamneză sau datele biografice relevante în raport cu scopul testării, observaţiile relevante reieşite în timpul desfăşurării examenului psihologic, interpretarea rezutatelor (sumarizarea) şi recomandările.

5. Dublarea laturii constatative cu cea formativă, cea diagnostică cu cea prognostică şi abordarea nomotetică cu cea idiografică (personalizarea examenului psihologic) face diferenţa între un raport bun şi unul slab, deşi aici elementul diferenţiator major este experienţa capitalizată de psiholog, competenţa lui de psihodiagnostician, ca şi lărgimea repertoriului psihodiagnostic de care dispune şi care-i permite să abordeze o gamă mai largă sau mai îngustă de probleme.

6. Aceasta înseamnă că nu aspectul formal (standard) dă valoarea unui raport psihologic, ci:

- gradul de acoperire a scopului şi a aşteptărilor clientului;

- măsura în care s-a operat cea mai potrivită selecţie a testelor şi a procedurilor disponibile la un moment dat, în raport cu solicitarea clientului;
- faptul de a include datele cele mai pertinente reieşite din observaţie, care au o mare valoare diagnostică, acestea fiind cele care transferă examinarea psihologică din zona nomoteticului în zona idiograficului;
- faptul dacă rezultatele sunt confirmate de timp (fidelitatea test-retest), de alţi evaluatori (fidelitatea interscoreri sau interspecialişti), de faptele anticipate (validitatea relativă la criteriu) sau de viaţa însăşi, aflată dincolo de pereţii laboratorului (validitatea ecologică);
- faptul de a oferi altor specialişti (profesori, medici, sociologi etc.) puncte de sprijin solide în deciziile sau în intervenţiile lor de specialitate;
- faptul de a ţine pasul cu cele mai noi şi avansate standarde din domeniu, furnizate de ţările cu mare tradiţie în testarea psihologică.

11.5. Activităţi propuse

11.5.1. Întocmirea unui raport de examinare psihologică

Raportul psihologic materializează şi finalizează o investigaţie psihologică de tipul testării/ evaluării, adică un examen psihologic. Mărimea şi gradul de elaborare al acestuia depind simultan de scopul testării/evaluării, de cerinţele expres formulate de cel care a comandat-o, de exigenţa şi profesionalismul psihologului. Chiar şi pentru testări foarte scurte trebuie să existe un raport, adică ceva care traduce datele cantitative în date calitative cu sens, inteligibile pentru client.

11.5.2. Structura raportului psihologic

Un raport de evaluare a funcţiilor intelectuale poate diferi semnificativ de unul care se referă la structuri şi funcţii de personalitate. Astfel, testul CPI, cu multitudinea sa de scale originare (18), la care se adaugă cele suplimentare (cel puţin tot atâtea) poate genera atât efort în sinteza şi de intercorelare încât, mai nou, această sarcină a fost preluată de computer (care nu poate suplini totuşi responsabilitatea psihologului, asumată prin semnătură). Aceste rapoarte computerizate au de regulă o parte non-narativă - raportul scorurilor brute şi standard, incluzând testele de semnificaţie statistică şi intervalele de încredere pentru scorurile de la test, şi una narativă. "Raporturile descriptive merg cu un pas dincolo de

raporturile de scoruri prin includerea unei interpretări a rezultatelor scală cu scală."[29] Lungimea raportului nu dă însă şi valoarea acestuia: profilurile furnizate de softurile contemporane fiind excesiv de detaliate, clinicianul trebuie să parcurgă rapoarte de peste zece pagini pentru a putea extrage câteva caracteristici definitorii pentru clientul său. Concluzia lui Gregory ni se pare una foarte valabilă: "Un raport descriptiv poate da o informaţie nepreţuită pe o jumătate de pagină."[30]

Dacă există totuşi diferenţe mari de format între raportul psihologic al funcţiilor intelectuale şi al profilului de personalitate, înseamnă că ne vom opri asupra celui cu structura mai simplă (funcţii intelectuale), deşi în practica curentă psihologul poate construi profile hipercomplexe (care include inteligenţa, memoria, motricitatea, atenţia şi personalitatea), atunci când vrea să surprindă integral sau multiaxial persoana.

Structura raportului psihologic va trebui să cuprindă:[31]

I. ***Date factuale*** şi de identificare ale clientului.

II. ***Întrebarea de referinţă*** (scopul) evaluării psihologice.

III. ***Testele administrate***, cu indicarea rezultatelor în note brute şi note standard, ca şi a scorului final (QI, Indice de Memorie, Indice de Motricitate).

IV. ***Datele de anamneză*** (pacienţi clinici) sau istoria personală a cazului.

V. ***Observaţiile relevante*** reieşite pe parcursul examinării.

VI. ***Interpretarea rezultatelor*** la teste prin raportarea datelor brute la etaloane, raportarea unora la altele, a tuturor la o teorie, la un set de ipoteze sau la fapte).

VII. ***Sumarizarea concluziilor.***

VIII. ***Recomandări***.

Mai detaliat, această secvenţă constă din următoarele opt capitole ale raportului, fiecare cu o scurtă definiţie, cu indicarea funcţiilor implicate, cu precizarea modului de realizare corectă şi a erorilor cele mai frecvente, dar şi a modului lor de evitare. În final se propune un exerciţiu imaginar pe un blanc anume elaborat.

Notă: raportul psihologic se parafează, se datează şi se semnează, pentru că din momentul emiterii lui devine un document oficial, cu consecinţe posibile asupra clientului, el putând fi invocat în instanţă sau în luarea unor decizii cu privire la persoana în cauză. Spre

[29] Gregory, R.J. (1996). *Psychological Testing. History, Principles and Applications*, second edition. Boston, London, Toronto etc: Allyn and Bacon, p. 575.
[30] Ibidem.
[31] Prelucrare după Richard, 1988.

deosebire de raportul psihologic computerizat, raportul psihologic presupune deci asumarea responsabilităţii prin semnătură şi parafă.

I. Datele factuale şi de identificare a subiectului

Definire: sunt datele care descriu caracteristicile demografice principale pentru identificarea subiectului şi includerea sa într-o categorie, cum ar fi: data naşterii şi data examinării, ce permit determinarea vârstei (în ani şi luni, căci la vârsta copilăriei dinamica dezvoltării este una foarte rapidă), mediul (rural/urban = R/U), sexul (Masculin/Feminin = M/F), studiile părinţilor copiilor examinaţi (generale, medii, superioare = G, M, S) sau ale celui în cauză, şcoala şi clasa sau facultatea (dacă este cazul), ultima şcoală absolvită (dacă este adult), mărimea şi poziţia sa în fratrie, familia de apartenenţă (căsătorit, divorţat, văduv, concubinaj) sau cea proprie (stare maritală = căsătorit/necăsătorit), ocupaţie, vechime, calificare etc. Observăm că pentru copii se pune problema *background*-ului cultural-economic şi educaţional; când cel examinat este un adult, aceste elemente se transferă de pe familie pe clientul însuşi, de care ne mai putem interesa privitor la starea locativă, indicele de aglomerare (număr de persoane pe număr de camere), venit (global sau pe membru de familie). Toate acestea lungesc proporţional durata şedinţei, de aceea trebuie să găsim un echilibru rezonabil între lungimea datelor completate şi relevanţa lor.

Funcţiile acestui capitol sunt următoarele:

- fixează datele cele mai relevante, care permit identificarea şi raportarea clientului la o clasă;
- datele factuale sunt variabile "naturale" (uneori "etichetă") în prelucrarea computerizată a seturilor de date similare;
- din prelucrarea bazei de date acumulate în timp se pot extrage sub-eşantioane mai mici, pentru a genera bareme/etaloane mai fine sau pentru a desprinde regularităţi tipice (prin lucrări de cercetare ştiinţifică).

Realizare corectă

- Dacă este structurată sub forma unei fişe cu o rubricatură adecvată, datele sunt uşor de strâns, putând avea o multitudine de utilităţi (vezi funcţiile).
- Deoarece raportul psihologic selectează doar datele (adevărate) din fişa de test (mai completă, mai tehnică şi mai analitică), formularea din fişă este diferită de cea din raport. În primul caz, se completează datele în rubrică, în al doilea li se dă o structură narativă sintetică.

De exemplu, datele de mai jos din fişa psihologică, ce rămâne la examinator:
Numele: Ionescu Mircea; ***Data naşterii***: 4. 01. 1975; ***Profesia***: sudor; ***Sex***: masculin (M)
Data examinării: 21 05 2005; ***Vârsta***: (în ani şi luni): 46 de ani şi 3 luni; ***Stare civilă***: căsătorit, 3 copii, ar putea arăta astfel în raportul psihologic: „Ionescu Mircea, de 46 ani şi 3 luni, de profesie sudor, cu studii medii (şcoala profesională plus liceul la seral), căsătorit, tatăl a 3 copii etc.

Erori probabile

- lungirea nejustificată a numărului de date de identificare solicitate clientului;
- contaminarea capitolului cu date rezultate din istoria personală, din observarea directă sau din testarea efectivă a clientului.

II. Scopul evaluării psihologice (întrebarea de referinţă)

Definire: această rubrică trebuie să indice fără dubiu cine a cerut evaluarea psihologică şi cu ce scop.

Funcţie: scopul ghidează întreaga evaluare psihologică ca număr şi tip de teste selecţionate, ca profunzime a investigarii, bogăţie a concluziilor şi nuanţare a recomandărilor. Dacă persoana care cere evaluarea nu poate preciza clar scopul, este datoria psihologului să-l reformuleze astfel încât el să devină unul explicit. Realizarea corectă va sintetiza într-o manieră cât se poate de clară şi concisă *de ce* şi *pentru cine* se face evaluarea psihologică.

Greşeli de evitat

- un scop „ciorchine”, amalgamând o multitudine de elemente, dintre care unele reieşite din anamneză;
- scopuri vagi sau prea generale („dorim o evaluare de ansamblu”, „ne interesează cum gândeşte”, „copilul nu are stare” etc.), care vor fi înlocuite cu scopuri specifice (consecinţele cognitive ale unei naşteri hipoxice, hiperkinezie cu deficit de atenţie, examen în vederea orientării şi ghidării vocaţionale etc.).

III. Testele administrate şi rezultatele obţinute (note brute şi note standard)

Descriere: pentru anumite categorii de teste pot exista foi de înregistrare / centralizare şi asociere a datelor pre-elaborate, în care se folosesc aceleaşi baterii de teste. De exemplu, pentru orientarea şcolară şi ghidarea vocaţională bateria va cuprinde teste de aptitudini (generale şi speciale), teste de memorie, un test de personalitate (HSPQ) şi teste de interese (Strong şi/sau Holland). În consecinţă şi pe foaia de raport psihologic ele vor fi trecute ca

atare, cu o coloană pentru note brute şi una pentru note standard, utile pentru a verifica informaţia de specialitate cînd raportul circulă printre psihologi sau alţi specialişti.

Funcţie: este aceea de a avea un sistem de control extern al elementelor pe care se bazează interpretările calitative care urmează, pentru a vedea în ce măsura testarea psihologică este adecvată scopului său declarat.

Greşeli de evitat: adăugarea sau omiterea de teste care nu au fost efectiv utilizate.

IV. Anamneza sau istoria personală a cazului

Descriere: anamneza este dictată de scopul explicit formulat al testării, investigaţia trecutului clientului trebuind să surprindă elementele semnificative din viaţa persoanei în raport cu acest fapt. De exemplu, dacă copilul are hiperkinezie cu deficit de atenţie (ADHD), vom căuta în istoria sa timpurie evenimente ce ar fi putut genera disfuncţia cerebrală minimă, adică o hipoxie la naştere, un Rh negativ, o circulară de cordon, o naştere provocată sau dificilă. Mai târziu ne vom interesa de traumatisme craniocerebrale, de asfixii mecanice sau cu monoxid de carbon; vom cerceta dacă semnele ADHD sunt centrate mai mult pe atenţie sau mai mult pe scăderea autocontrolului (impulsivitate), dacă fenomenologia a diminuat odată cu vârsta, dacă deficitul a avut concomitenţe legate de funcţia grafică (desen, scris, citit), încercând să distingem între un temperament necontrolat, instabil (coleric) şi sindromul amintit. Pot fi cercetate şi antecedentele heredo-colaterale, pentru a diferenţia între aspectul dobândit şi cel ereditar al tulburării.

Funcţie: comportamentul actual al subiectului are o istorie care poate fi reconstituită mai larg sau mai nuanţat, în funcţie şi de calitatea surselor de informaţie disponibile. *Realizare corectă:* ca şi în cazurile de informaţii obţinute prin spionaj, se pune problema calităţii acestora, care depinde de calitatea sursei, dar şi de posibilitatea coroborării surselor. Astfel, mamele dau informaţii mult mai detaliate şi acurate decât taţii, părinţii decât bunicii, rudele decât persoanele neutre etc. Când există şi o electroencefalogramă, expertiza altor specialişti (date consemnate în fişa medicală), diagnosticele sunt mai certe şi căutarea este mult uşurată. Specialiştii atrag atenţia că trebuie înregistrate datele comportamentale observate şi interpretarea dată acestora. De exemplu, este preferabil să notăm că „elevul umblă în timpul orelor printre bănci” decât că „este agitat şi fără stare” sau că „este indisciplinat”, „nu are răbdare”, „tulbură clasa” etc. Aceasta deoarece persoane diferite atribuie semnificaţii diferite, în contexte diferite, aceloraşi fapte de conduită, urmând a evita operaţionalizările incorecte, mai ales când este vorba de surse nespecializate.

Greşeli de evitat

Sondarea minuţioasă a trecutului unei persoane supraîncarcă rubrica respectivă cu o multitudine de date irelevante în raport cu scopul investigaţiei, după cum una prea "de la depărtare" omite chiar elemente dintre cele mai semnificative. Mai jos dăm exemplu de elemente semnificative din anamneza copiilor cu dificultăţi şcolare sau cu probleme de dezvoltare care pot fi incluse în raportul psihologic.

- Probleme antenatale, perinatale şi postnatale importante (de exemplu disgravidie, naştere precipitată sau înainte de termen, naştere pe uscat, cianoză la naştere, Rh negativ, prematuritate, imaturitate, circulară de cordon, asfixie albastră sau albă etc.).
- Boli importante ale micii copilării.
- Retarduri legate de vorbit, mers, control sfincterian (primul pas, primul cuvânt).
- Tulburări ale senzorialităţii (văz/auz) şi gravitatea lor.
- Lateralizare, dominanţă inversă sau încrucişată.
- Evoluţia timpurie a limbajului, a comunicării şi socializării.
- Apetitul pentru joc.
- Adaptabilitatea la colectivitate (creşă, grădiniţă) şi uşurinţa integrării.
- Evenimentele majore din viaţa copilului sau a familiei sale (separaţii, divorţuri etc.).
- Date de climat familial (relaţii cu fraţii, bunicii, rudele, prietenii).
- Rezultatele şcolare actuale şi evoluţia lor în timp.

Este inutil să spunem că altă categorie de probleme (tumori, accidente neurologice, devianţă sau delincvenţă juvenilă, tulburări instrumentale, deficienţe caracteriale) au alte „grile de anamneză", care se structurează prin acumularea de experienţă, dar şi prin parcurgerea literaturii de specialitate. Cu atât mai mult examenul adultului sau investigaţii mai speciale (avizarea pe post, de exemplu) au alte tipuri de *screening* al istoriei personale. Este de asemenea foarte important de spus că cea mai mare parte din datele consemnate rămân în fişa de examinare (cea care face sinteza datelor), în raport trecându-se datele filtrate şi cu adevărat concludente în raport cu problema care face obiectul investigaţiei.

V. Observaţiile relevante din timpul examinării

Descriere: este locul în care diagnosticul medical şi cel psihologic se suprapun în bună măsură. Şi psihologul trebuie să dezvolte un ascuţit simţ clinic, mai ales în examinările individuale, care furnizează acea informaţie suplimentară, nicicând surprinsă de teste şi care se referă la:

- Conduita generală în timpul examenului.
- Starea de mobilizare atenţională şi fluctuaţiile atenţiei datorate scăderii motivaţiei, interesului sau epuizării resurselor energetice.
- Susţinerea proceselor reglatorii şi energetice (autocontrol, impulsivitate, declin energetic).
- Stilul de abordare cognitivă, strategiile utilizate (mai ales la probele de performanţă).
- Comentariile cu care se asociază reuşita/nereuşita, toleranţa la frustrare, reacţia la eşec, susţinerea mnezică, promptitudinea reacţiilor.

Este evident că în orice testare a aptitudinilor (şi a inteligenţei în special) persoana se manifestă ca întreg, lucru precizat şi de Binet, Piaget sau Wechsler, care au căutat să deriveze concluzii despre personalitatea subiectului din examenul inteligenţei. Piaget a propus şi realizat standardizarea observaţiilor în timpul aplicării probelor sale cognitive. Psihologul obişnuit ştie că dacă a folosit cum trebuie observaţia, el poate „agrementa" raportul său cu acele elemente decisive ce fac diferenţa dintre examenul lipsit de suflet - computerizat - şi cel efectuat de o fiinţă umană.

Rezolvarea corectă: problema este aceea că din multitudinea datelor observate şi consemnate în fişa de examen să selectăm doar acele informaţii concludente, într-o formă explicită. De exemplu, în loc să trecem că „Ionel este inhibat, foarte dependent afectiv" (concepte vagi) putem utiliza formula „Ionel a vrut să fie examinat doar în prezenţa părinţilor săi, pe care-i solicita la cea mai mică neclaritate; la un moment dat el s-a ridicat de pe scaun şi nu a mai vrut să continue, refugiindu-se în spatele mamei, deşi Ionel are deja şapte ani şi jumătate. A comentat că şcoala nu este bună la nimic, poate doar să-i chinuie în zadar pe copii. A revenit în examen după insistenţe şi rugăminţi, arătându-se foarte îngrijorat la tot ce ar fi putut însemna eşec sau nereuşită. La comunicarea rezultatelor a ţinut neapărat să fie de faţă, cerând explicaţii şi întrebând dacă ce a făcut el este bine. La plecare nu şi-a luat la revedere şi tata a comentat că este foarte puţin probabil să mai vină din nou la examen, pentru că examinarea seamănă prea mult cu şcoala."

Erori posibile

- Contaminarea reciprocă a datelor furnizate de observaţie şi a celor de anamneză.
- Selectarea datelor de observaţie care „se potrivesc" cu diagnosticul psihologic sau psihiatric anterior formulate.

- Contaminarea cu alte surse de informaţii (de la învaţătoare, profesori, medici, asistenţi sociali).

VI. Interpretarea rezultatelor

Definire: este cea mai complexă şi cea mai importantă parte a raportului psihologic deoarece face sinteza datelor psihometrice cu cele de anamneză, de observaţie şi provenite din surse colaterale, cu scopul a le da un sens inteligibil pentru părinţi, client, beneficiar, într-o manieră obiectivă, dar care nu lezează, informaţia fiind explicită, utilizabilă şi în alte contexte.

Funcţii

- Valorizează datele obţinute în vederea atingerii scopului examinării.
- Dă elemente clare beneficiarului imediat şi celor îndepărtaţi (angajatorul, profesorul, medicul, autoritatea locală, poliţia etc.).
- Face ca datele să conveargă spre o finalizare, spre o concluzie.

Realizarea corectă nu se sprijină pe o reţetă. În psihologia dezvoltării, de exemplu, această etapă face dovada experienţei, competenţei, adâncimii şi pluridisciplinarităţii în pregătirea unui psiholog. Dacă părinţii suspectează de retard mintal o fată mai lentă, cu hidrocefalie în antecedente, ei au nevoie de un sfat avizat, care le-ar putea motiva îngrijorarea sau „adormi vigilenţa". Problema este ce spui şi cât spui pentru ca bunăstarea clientului să fie servită, fără a omite totuşi adevărul. Tendinţa ultimilor ani în America este aceea de a fi mai aproape de adevăr, chiar şi pentru persoane cu probleme reale, care-şi pot mobiliza mai bine resursele adaptative pentru a înfrunta un deficit sau handicap.

Interpretarea trebuie să fie *corectă*, *completă* şi *obiectivă*. Trebuie să sublinieze minusurile, dar şi plusurile, dacă se poate după stabilirea de intercorelaţii între toate datele sau după conturarea unei scheme explicative ordonatoare.

Erori evitabile

- *Suprasimplificarea*: interpretarea de genul: „rezultatul la testul cuburilor este peste medie, deoarece subiectul a obţinut nota standard 17". În fond, putem porni de la o privire de ansamblu a rezultatelor, să vedem gradul lor de omogenitate/heterogenitate, dominantele etc. De exemplu, cota standard 14 la cuburi poate să însemne coordonare oculo-manuală foarte bună, susţinută de un tempo de lucru alert (deoarece a primit bonificaţii de timp însemnate), iar la testul Cod B - probă de învăţare asociativă -

subiectul a obţinut de asemenea un scor foarte ridicat, cota standard 16. Deoarece şi testul Bender-Gestalt indică o performanţă superioară, se poate concluziona că un factor major al inteligenţei - cel spaţial - este supradezvoltat, avem deci de-a face aici cu premisele unei aptitudini ce trebuie valorificată. Pe de altă parte, testul de vocabular, memoria verbală şi memoria ritmurilor se află la un nivel mediocru, ceea ce sugerează o puternică asimetrie între factorul verbal şi cel spaţial, tipică stângăciei. Aceasta ar putea explica eşecurile elevului la limbi, biologie şi istorie. Bâlbâiala de la 3 ani, instalarea tardivă a dominanţei (după 7 ani, odată cu şcoala, dar cu tendinţe remanente spre ambidextrie) susţin ipoteza unei organizări corticale particulare, cu o dominanţă emisferică vagă."

- *Interpretările incorecte* provenite din necunoaşterea testelor (a instrumentelor psihometrice), din insuficienta cunoaştere a teoriei psihologice (de unde nevoia informării / formării continue) sau din supradimensionarea unei singure informaţii. De exemplu, un tânăr psiholog de la Laboratorul de sănatate mintală dădea următoarea formulare: „dacă nu există altă contraopinie, pacientul X este suspect de schizofrenie", şi aceasta după un singur test proiectiv, Testul Arborelui! Confuzii de acelaşi tip sunt şi cele care concluzionează în legatură cu excesul de energie al hiperkineticului (care nu debordează de energie, cum se creden mod curent, ci doar dă impresia, din cauza consumului haotic al acesteia).
- Uneori psihologul *se rezumă doar la scopul testării,* ferindu-se să valorifice date valoroase care-l depăşesc ca putere explicativă, alteori examenul nu ţine deloc cont de scopul de referinţă, mergând pe acelaşi şablon, stereotip, indiferent de tipologia solicitărilor întâlnite.

VII. Sumarizarea concluziilor

Definire: în această secţiune a raportului psihologic se rezumă în unul-două paragrafe rezultatele de ansamblu ale întregului examen.

Funcţii

- Aceea de a răspunde precis şi concis la întrebarea de referinţă care a declanşat evaluarea psihologică.
- Aceea de a face legătura cu ultima secvenţă a raportului, cea a propunerilor şi recomandărilor.

Greşeli: prolixitatea şi redundanţa.

VIII. Recomandări

Definire: sugestiile oferite sunt date în sensul potenţării calităţilor insuficient puse în valoare de către subiect şi a optimizării sau compensării aspectelor deficitare.

Funcţii: aceea de a depăşi planul constatativ şi de a oferi soluţii problemelor diagnosticate, de a lărgi cadrul de referinţă prin implicarea altor specialişti (dacă e cazul) şi de a găsi soluţii problemelor pe care subiectul le are fără ca acestea să fi fost în mod expres obiectul investigaţiei.

Greşeli de evitat

- Recomandările să fie pertinente, la obiect şi fezabile.
- Ele să rezulte din datele de investigaţie, să fie solid susţinute de datele empirice, dar şi de teoria din domeniu.
- Să permită constatarea efectelor ameliorative după o perioadă de timp (prin retest).

Exerciţiu pentru portofoliul de evaluare

Realizaţi un *Raport psihologic* pentru un caz imaginar, respectând paşii şi recomandările corespunzătoare.

BIBLIOGRAFIE

1. Aiken, L.R. (1997). *Psychological Testing and Assessment*. 9th ed. Boston, London, Toronto, Sydney, Tokyo, Singapore: Allyn and Bacon.
2. Albu, M. (1998). *Construirea şi utilizarea testelor psihologice*. Cluj-Napoca: Editura „Clusium".
3. Allport, G.W. (1991). *Structura şi dezvoltarea personalităţii*. Bucureşti: Editura Didactică şi Pedagogică.
4. Anastasi, A. (1974). *Psychological Testing*. New York: Macmillan Publishing Co.
5. Băban, A., Derevenco, P., Eysenck, S.B.G. (1990). Testul E.P.Q. În: *Revista de psihologie*, 1, 37-45.
6. Băiceanu, L. (2004). *Dicţionar ilustrat de psihologie englez – român*. Bucureşti: Editura tehnică.
7. Bontilă, Gh. (1971). *Aptitudinile şi măsurarea lor*. Bucureşti: Centrul de Documentare şi Publicaţii al Ministerului Muncii.
8. Carroll, J.B. (1993). *Human Cognitive Abilities*. Cambridge: Cambridge University Press.
9. Constantin, T. (2004). *Evaluarea psihologică a personalului*. Iaşi: Editura Polirom.
10. Bogáthy, Z. (1978). *Îndreptarul psihologului industrial*. Bucureşti: Editura Ştiinţifică şi Enciclopedică.
11. Cattell, R.B. (1957), *Personality and Motivation. Structure and Measurement*. New York: World Book Co.
12. Cattell, R.B. (1960). *Manuel pour l'application de l'Echelle d'Anxieté*. Paris: Ed. C.P.A.
13. Cattell, R.B., Beloff, H. (1966). *Manuel d'application du H.S.P.Q.* Paris: Ed. C.P.A.
14. Clocotici, V., Stan, A. (2000). *Statistică aplicată în psihologie*. Iaşi: Editura Polirom.
15. Cohen, R.J., Swerdlik, M.E., Smith, D.K. (1992). *Psychological Testing and Assessment. An Introduction to Tests and Measurement*. second ed. California: Mayfield Publishing Company, Mountain View.
16. Cronbach, L.J. (1970). *Essential of Psychological Testing*. New York: Harper and Row.
17. Deri, S. (2000). *Introducere în testul Szondi*. Bucureşti: Editura Paideia.
18. Durand, G. (1999). *Aventurile Imaginii. Imaginaţia simbolică. Imaginarul*. Bucureşti: Editura Nemira.
19. Dworetzky, D.J., Davis, N.J. (1989). *Human development. A life span approach.* St Paul, New York: West Publishing Company.
20. Enăchescu, C. (1975). *Expresia plastică a personalităţii*. Bucureşti: Editura Ştiinţifică.

21. Enăchescu, C. (1977). *Psihologia activităţii patoplastice*. Bucureşti: Editura Ştiinţifică şi Enciclopedică.
22. Eysenck, H.J. (1953). *The Structure of Human Personality*. New York: Viley.
23. Eysenck, H.J. (1975). *The Measurement of Personality*. London: Routledge & Kegan.
24. Eysenck, H.J, Eysenck, S.B.G. (1975). *Manual of the Eysenck Personality Questionnaire*. London: University of London Press.
25. Eysenck, M.W., Keane, M.T. (1995). *Cognitive Psychology. A student's handbook*. 3rd edition, Hove East Sussex, UK: Psychology Press Ltd.
26. Gardner, H. (1985). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books
27. Gregory, R.J. (1994). *Psychological Testing. History, Principles and Applications*, second edition. Boston, London etc.: Alyn and Bacon.
28. Guilford, J.P. (1959). *The Nature of Human Intelligence*. New York: Mc Graw Hill Book Company.
29. Harris, D.B. (1963). *Children's drawings as measures of intellectual maturity*. New York: Harcourt, Brace & World.
30. Havârneanu, C. (2000). *Cunoaşterea psihologică a persoanei*. Iaşi: Editura Polirom.
31. Holban, I (1970). *Probleme de psihologia muncii*. Bucureşti: Editura Ştiinţifică.
32. Horghidan, V. (1992). *Metode de psihodiagnostic*. Bucureşti: Editura Didactică şi Pedagogică.
33. Kaplan, R.M., Saccuzzo, D.P. (1993). *Psychological Testing. Principles, Applications and Issues*, 3th ed. Belmont, California: Brooks Cole Publishing Company, Inc.
34. Kline, P. (1995). „La construction des tests". În: Beech, J.R., Harding, L., *Tests, mode d'emploi. Guide de psychométrie*. Paris: Ed. C.P.A.
35. Kulcsar (1976). Testul Raven şi variantele sale. În: *Îndrumător psidodiagnostic,* vol. II, Cluj-Napoca: Reprografia Univerităţii Babeş Bolyai.
36. Laveault, D., Grégoire, J. (2002). Itroduction aux théoriex des tests en Psychologie et en sceeces de l'éducation. 2e edition. Bruxelles: Éditions De Boeck Université, p. 199.
37. Meeker, M.N. (1969). The Structure of Intellect. Its interpretations and uses. Columbus, Ohio: Charles E. Merrill Company.
38. Meili, R. (1962). *Manuel du diagnostique psychologique*. Paris: P.U.F.
39. Minulescu, M. (1996). *Chestionarele de personalitate în evaluarea psihologică*. Bucureşti: Garell Publishing House.

40. Mitrofan, N. (1993). *Testarea psihologică a copilului mic*. Bucureşti: Editura Press Mihaela.
41. Mitrofan, N., Mitrofan, L. (2005). *Testarea psihologică. Inteligenţa şi aptitudinile.* Iaşi: Editura Polirom.
42. Mitrofan, N. (2001). Psihometria şi direcţiile ei de dezvoltare la început de mileniu. În M. Zlate (coord.), *Psihologia la răspântia mileniilor*. Iaşi: Editura Polirom.
43. Nunnally, J.C. (1978). *Psychometric Theory*. (2nd Ed.). New York: McGaw Hill.
44. Parot, F., Richelle, M. (1995). *Introducere în psihologie. Istoric şi metode*. Bucureşti: Editura Humanitas.
45. Perse, J. (1972). Collaboration du psychologie au dépistage et au diagnostique des maladies mentales. În: Reuchlin, M. (red.) *Traité de psychologie apliquée*. Paris, Press Univ. de France.
46. Piéron, H. (coord.) (2001). *Vocabular de psihologie*. Bucureşti: Editura Univers Enciclopedic.
47. Pitariu, H. (1983). *Psihologia selecţiei şi orientării profesionale*. Cluj-Napoca: Editura Dacia.
48. Pitariu H. (1974). Testul Domino 48 (D 48); Testul Domino 70 (D 70). În: *Îndrumător psihodiagnostic*, vol. II, Cluj-Napoca: Reprografia Universităţii Babeş Bolyai.
49. Popescu Neveanu, P. (1977). *Psihologie generală*. Tipografia Universităţii Bucureşti.
50. Radu, I. (coord.) (1991). *Metodologie psihologică de analiză a datelor*. Cluj-Napoca: Editura Sincron.
51. Reber, A.S. (1985). *Dictionary of Psychology*. London, New York etc.: Penguin Books.
52. Reuchlin, M. (1992). *Introduction à la recherche en psychologie*. Paris: Éditions Nathan.
53. Reuchlin, M. (2000). *Psihologie generală*. Bucureşti: Editura Ştiinţifică.
54. Ribault, C. (1965). Le dessin de la maison chez l'enfant. În: *Revue de Neuropsychiatrie Infantile***,** 13, nr. 1-2.
55. Rorschach, H (2000). *Manual de psihodiagnostic*. Bucureşti: Editura Trei.
56. Roşca, M. (1972). *Metode de psihodiagnostic*. Bucureşti: Editura Didactică şi Pedagogică.
57. Rozorea, A., Sterian, M. (2000). *Testul arborelui.*Bucureşti: Editura Paideia.
58. Stan, A. (2002). *Testul psihologic. Evoluţie, construcţie, aplicaţii*. Iaşi: Editura Polirom.
59. Sternberg, R.J. (1985). *Beyond IQ: A triarchic treory of human intelligence.* New York: Cambridge University Press.

60. Sternberg, R.J., ed. (1990). *Handbook of human intelligence*. Cambridge, New York etc.: Cambridge University Press.
61. Sternberg, R.J. (1993). Intelligence is more than IQ: The practical size of intelligence**.** In *Jurnal of Cooperative Education*, 28 (2), pp. 6-10.
62. Szamosközi, Şt. (1997). Psihometria clasică şi evaluarea formativă. În *Cogniţie, creier, comportament*, vol. 1, nr. 2, iunie, Cluj-Napoca.
63. Şchiopu, U. (1974). *Introducere în psihodiagnostic*. Bucureşti: Tipografia Universităţii.
64. Şchiopu, U. (coord.) (1997). *Dicţionar enciclopedic de psihologie*. Bucureşti: Editura Babel.
65. Ştefănescu-Goangă, Fl. (1946). *Măsurarea inteligenţei. Revizuirea, adaptarea şi completarea scării de inteligenţă Binet-Simon*. Cluj-Napoca: Editura Institutului de psihologie.
66. Witkin, H.A., Dyk, R.B., Faterson, H.F., Goodenough, D.R., Karp, S.A. (1962). *Psychological differentiation. Studies of development.* NewYork, London: John Wiley and Sons, Inc.
67. Zazzo, R., Gilly, M., Verba-Rad, M. (1968). *Nouvelle échelle métrique de l'intelligence*. Paris: Librairie Armand, Colin.
68. Zimmerman I.L., Woo-Sam, J.M. (1973). *Clinical Interpretation of the Wechsler Adult Intelligence Scale*. New York, San Francisco, London: Grune & Straton.
69. Zlate, M. (1999). *Psihologia mecanismelor cognitive*. Iaşi: Editura Polirom.

CUPRINS

Capitolul 1. Introducere …….. 3
1.1. Psihodiagnoza, domeniu de vocaţie al profesiunii de psiholog 3
1.2. Psihodiagnoză, testare şi evaluare 4
1.3. Marile domenii ale psihodiagnosticului 7
1.4. Probleme etice ale psihodiagnozei 11

Capitolul 2. Testul psihologic, instrumentul de bază al psihologiei aplicate 15
2.1. Scurt istoric al apariţiei şi dezvoltării testelor 15
2.2. Definirea testelor 16
2.2.1. Eşantionul de comportament 16
2.2.2. Standardizarea 19
2.2.3. Fidelitatea 19
2.2.4. Validitatea 20
2.3. Administrarea testelor 21
2.3.1. Îndatoririle examinatorului înainte de administrarea testelor..... 22
2.3.2. Faza de administrare a testelor 23
2.3.3. Faza post-test 25
2.4. Examinatorul şi variabilele situaţionale 25

Capitolul 3. Dificultatea şi sensibilitatea unui test psihologic sau educaţional 26
3.1. Dificultatea unui test 26
3.2. Sensibilitatea unui test 27
3.3. Construcţia unui test psihologic ……... 31

Capitolul 4. Fidelitatea testelor 33
4.1. Problematica generală a fidelităţii testelor 33
4.2. Fidelitatea test-retest …… 34
4.3. Coeficientul formelor paralele ……. 34
4.4. Coeficienţii de consistenţă internă ……... 35
4.5. Factorii care intervin în determinarea fidelităţii 38

Capitolul 5. Factori care afectează fidelitatea testelor 40
5.1. Întinderea diferenţelor individuale 40
5.2. Lungimea unui test ……..... 41
5.3. Dificultatea testului …….. 42
5.4. Testele cu limită de timp …….. 42

5.5. Fidelitatea şi eroarea de măsurătoare 43
5.5.1. Eroarea standard a măsurătorii (SEM) 43
5.5.2. Eroarea tip a estimării 45
5.5.3. Interpretarea fidelităţii 46
5.5.4. Generalizabilitatea 46
5.5.5. Analiza de varianţă şi planul de optimizare 48

Capitolul 6. Validitatea 49
6.1. Validitatea de faţadă 49
6.2. Validitatea relativă la conţinut 50
6.3. Validitatea relativă la criteriu 50
6.4. Validitatea relativă la construct 52

Capitolul 7. Alte feţe ale validităţii 54
7.1. Validitatea relativă la construct (conceptuală) 54
7.1.1. Metodele corelaţionale simple 56
7.1.2. Matricile multitrăsătură-multimetodă 56
7.1.3. Studiul trăsăturilor latente 58
7.2. Validitatea diferenţială 60
7.3. Teoria deciziei aplicată la testele psihologice, asociată validităţii predictive... 61
7.4. Precauţii legate de evaluarea coeficienţilor de validitate predictivă 63
7.5. Interpretarea coeficienţilor de validitate 64

Capitolul 8. Analiza de itemi 66
8.1. Precizări terminologice 66
8.2. Dificultatea itemilor 68
8.3. Discriminabilitatea itemilor 75
8.4. Raportul dintre dificultatea şi discriminabilitatea itemului 76
8.5. Indici corelaţionali ai discriminării itemilor 79
8.6. Curba caracteristică a itemului (CCI) 81

Capitolul 9. Normele şi etalonarea testelor. Clasificarea testelor 83
9.1. Normele testelor 83
9.2. Etalonarea testelor 83
9.3. Clasificarea testelor 90
9.3.1. Clasificarea testelor după modul de administrare 91
9.3.2. Clasificarea testelor după timpul de execuţie 91
9.3.4. Clasificarea testelor după modul de executare a sarcinii de către subiect 92
9.3.4. Clasificarea testelor după modul de procesare implicat 93

9.3.5. Clasificarea testelor după constanţa conţinutului de la o administrare la alta 96
9.3.6. Clasificarea testelor după modul de cotare 96
9.3.7. Clasificarea testelor după modul de interpretare a scorurilor 96
9.3.8. Clasificarea testelor după numărul variabilelor implicate 97
9.3.9. Clasificarea testelor din punctul de vedere al procesului psihic investigat 97

Capitolul 10. Dezvoltarea scalelor psihologice 98
10.1. Introducere 98
10.2. Paşii de urmat în dezvoltarea scalelor 98
10.2.1. Pasul întâi: determinăm cu claritate ceea ce dorim să măsurăm 98
10.2.2. Pasul al doilea: generarea unui set mare de itemi 103
10.2.3. Pasul al treilea: determinarea tipului de format al măsurătorii 108
10.2.4. Pasul al patrulea: revizia setului iniţial de itemi de către experţi ... 117
10.2.5. Pasul al cincilea: administrarea setului iniţial de itemi 118
10.2.6. Pasul al cincilea: evaluarea itemilor din setul iniţial 118
10.3. Criterii de selecţie a itemilor pentru oprimizarea lungimii scalei 127
10.4. Alte analize utile pentru determinarea fidelităţii scalei 130
10.5. Proiect aplicativ:
construirea unui chestionar pentru o trăsătură de personalitate 132

Capitolul 11. Raportul psihologic 135
11.1. Obiective 135
11.2. Lecturi recomandate pentru acest curs 135
11.3. Materiale şi suporturi necesare 135
11.4. Idei de bază pentru întocmirea unui raport psihologic 136
11.5. Activităţi propuse 137
11.5.1. Întocmirea unui raport de examinare psihologică 137
11.5.2. Structura raportului psihologic 137

Bibliografie 147

Cuprins 151